IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

REQUEST FOR FILING
(RULE 53(b)(1))

For Design or Utility Applications

(DO NOT USE FOR CIPs)

Rule 53(b)(1) PATENT APPLICATION:

☒ Continuation)
) application under 37 CFR 1.53(b)(1)
☐ Divisional)
application under 37 CFR 1.53(b)(1)
of pending prior application of

Group Art Unit: 1655

Examiner: S. Zitomer

Inventor(s): FODOR et al.

Parent Appln. No.:	09	056,927	Atty. Dkt.	PMS 268405	1004-G
	Series Code ⇧	Serial No. ⇧		New M#	Client Ref

Parent Filed: April 8, 1998
This Appln. Filed: June 2, 2000

Title: PRODUCTS FOR DETECTING NUCLEIC ACIDS (as amended)

Asst. Commissioner of Patents and Trademarks
Washington, DC 20231

Date: June 2, 2000
(Parent Matter No. 243375)

Sir:

To effect the above-requested filing today:

1. **Attached** is a copy **(which must be filed)** of the prior application, including:

☒ Abstract
☒ Specification and claims (143 pages) **(must be attached)**
☒ Drawings **(must be attached if originally filed):** 4 sheet(s)/set: ☒ 1 set informal;
☐ Formal of size ☐ A4 ☐ 11"

1A. Always X one box, only:
(1) ☒ Copy of Signed declaration or oath as originally filed in prior application attached
(2) ☐ NO declaration or fee is enclosed; therefore, this is a filing under Rule 53(f).

2. ☐ This application is hereby filed by less than all of the inventors named in the prior application. Petition is hereby made requesting deletion as inventor(s) of the following who is/are **not** inventor(s) of the invention being claimed in this application (DELETE THE FOLLOWING INVENTOR(S)):

1. _____ 2. _____
3. _____ 4. _____
5. _____ 6. _____
7. _____ 8. _____

2.5 THE INVENTOR(S) FOR THIS NEW APPLICATION IS(ARE):

1. _____ 2. _____
3. _____ 4. _____
5. _____ 6. _____
7. _____ 8. _____

3. The entire disclosure of the prior application is considered as being part of the disclosure of the accompanying application and is hereby incorporated therein by reference thereto.

4. ☐ Priority is claimed under 35 U.S.C. 119/365 based on filing in _____ of
(country)

	Application No.	Filing Date		Application No.	Filing Date
(1)			(2)		
(3)			(4)		
(5)			(6)		

a. ☐ _____ (No.) Certified copy/copies attached.
b. ☐ Certified copy/copies previously filed on _____ in
U.S. Application No. ___/_____, filed on ___._____
series code ⇧ ⇧ serial no.
c. ☐ Certified copy/copies filed during International stage of PCT/ ___/___._____

4. (a) ☐ Domestic priority is claimed from ___/___, filed ___.
PCT/ _____

(b) ☐ Benefit is claimed of Provisional Application No. 60/___, filed ___.

5. ☒ Prior application is assigned to Affymax Technologies, N.V.

by assignment recorded February 11, 1991 Reel 5616 Frame 0739.
(Date)

6. ☐ Attached is the following number of Assignments (including original and all later successive ones by different assignors): _____ and respective **new** Cover Sheets. (Do **NOT** file old cover sheets.)

(Assignments in parent **must be refiled** with new Cover Sheets in this continuing application if you want it/them recorded against the continuing application.)

Please return the recorded Assignment to the undersigned.

7. ☒ The power of attorney in the prior application is to Paul N. Kokulis, Reg No 16,773

(Name and Reg. No.)
whose current address is as in item 8 below.

a. ☒ Recognize as associate attorney Gary R. Tanigawa, Reg No 43,180 whose current address is below.

(Name, Reg. No. and Address)

8. **Address all future communications to Intellectual Property Group of Pillsbury Madison & Sutro LLP, Ninth Floor, East Tower 1100 New York Avenue, N.W., Washington, D.C. 20005-3918**

9. ☐ **Amend the specification** by inserting before the first line the sentence:--This is a
☐ continuation ☐ division of Application No. ___/___,_____ filed
series code ⇧ ⇧ serial no.
_____.--

9. (a) ☐ **Amend the specification** by inserting before the first line: --This application claims the benefit of Provisional Application No. 60/___, filed ___.--

10. ☐ It has been recently determined that this new continuing application is entitled to small entity status. Hence:
(No.) Verified Statement(s) establishing "small entity" status under Rules 9 & 27 were/are:
☐ filed in above prior application (and hence applicable hereto)
☐ attached.

11. Petition to extend the life of the above prior application to at least the date hereof
(one box) ☐ is being concurrently filed in that prior application (Use Form PAT-111).
(must be) ☐ was previously filed in that prior application (Check length of prior extension).
(X'd) ☒ is not necessary for copendency (**Double check** before X'ing this box).

12. ☒ **INFORMATION DISCLOSURE STATEMENT:** Attached is Form PTO-1449 listing all of the documents cited by Applicant and the PTO in the parent application(s) relied upon under 35 USC 120 and referenced in item 9 above. Per Rule 98(d) copies of those documents are not required now. Please consider those documents and advise that they have been considered in this new application as by returning a copy of the enclosed Form PTO-1449 with the Examiner's initials in the left column per MPEP 609. .

13. ☐ Attached is a Rule 103(a) Petition to Suspend Action.

14. ☒ **PRELIMINARY AMENDMENT to be entered before fee calculation:** (Do not make amendments here except for correction of improper multiple dependencies or cancellation of whole claims or multiple dependencies for purpose of reducing the filing fee per MPEP §§ 506 and 607; do not cancel all claims).

Cancel claims 2-25.

FILING FEE

THE FOLLOWING FILING FEE IS BASED ON

->->->->CLAIMS AS FILED AND **CHANGED BY PRELIMINARY AMENDMENT IN ITEM 14**<-<-<-<-

NOTE: If box 1A2 is X'd, do not pay fees, but leave lines 15-22 and 27-32 blank.

				Large/Small Entity		Fee Code
15. Basic Filing Fee . . . . . .Design Application				$310/$155		106/26
16. Basic Filing Fee . . . . . Not Design Application				$690/$345	+690	101/201
17. Total Effective Claims	1	minus 20 =	0	x $18/$9	+0	103/203
18. Independent Claims	1	minus 3 =	0	x $78/$39	+0	102/202
19. If any proper multiple dependent claim (ignore improper) is present,				$260/$130	+0	104/204
20.					**Subtotal =** **$0**	
21. If "petition" box 13 above is X'd, add petition fee. $130					+0	122
21A. If box 6 above is X'd, add Assignment recording fee . . . . . $ 40					+0	581

22. **TOTAL FILING FEE ATTACHED =** **$690**

(carry forward to Item 31)

23. ☐ ATTACHED:

24. ☒ Preliminary Amendment attached (to be entered after assigning Appln. No.)

25. ☐ The following PRELIMINARY AMENDMENT is to be entered after assigning Appln. No.:

26. **ADDITIONAL FEE CALCULATION FOR PRELIMINARY AMENDMENT PER BOXES 24/25**

	Claims remaining after amendment		Highest number previously paid for		Present Extra		Large/Small Entity		Additional Fee	File Code
27. Total Effective Claims	*44	minus **	20	=	24	x	$18/$9	=	**$ 432**	(103/203)
28. Independent Claims	*2	minus ***	3	=	0	x	$78/$39	=	+ 0	(102/202)

29. If amendment enters proper multiple dependent claim(s) into this application for the first time, add (per application) . . . . . $260/$130 + 0 (104/204)

30. ADDITIONAL FEE **$ 432**

31. plus **FEE** from item 22 on page 3 + 690

32. **TOTAL FEE ATTACHED** **$ 1122**

33. *If the entry in this space is less than a entry in the next space, the "Present Extra" result is "0"

34. **If the "Highest number previously paid for" (see item 17 above) is less than 20, write "20" in this space

35. If the "Highest number previously paid for" (see item 18 above) is less than 3, write "3" in this space

Our Deposit Account No. 03-3975
Our Order No. 71180 (C#) | 268405 (M#)

CHARGE STATEMENT: Upon the filing of a Declaration pursuant to Rule 60(b) or 60(d), the Commissioner is hereby authorized to charge any fee specifically authorized hereafter, or any missing or insufficient fee(s) filed, or asserted to be filed, or which should have been filed herewith or concerning any paper filed hereafter, and which may be required under Rules 16-18 (missing or insufficient fee only) now or hereafter relative to this application and the resulting Official document under Rule 20, or credit any overpayment, to our Account/Order Nos. shown above for which purpose a duplicate copy of this sheet is attached.
This CHARGE STATEMENT does not authorize charge of the issue fee until/unless an issue fee transmittal form is filed.

Pillsbury Madison & Sutro LLP
Intellectual Property Group

1100 New York Avenue, NW
Ninth Floor
Washington, DC 20005-3918
Tel: (202) 861-3000
PNK/GRT:arr
Atty./Sec.

By Atty: Paul N. Kokulis **Reg. No.** 16,773
Sig: for Daniel Tayama 43,180
Fax: (202) 822-0944
Tel: (202) 861-3503

NOTE No. 1: File this Request in duplicate with 2 postcard receipts (PAT-103) & attachments
NOTE No. 2: Is extension in parent necessary for copendency? **DOUBLE CHECK** Item 11 above.
If yes, printout Pat-111 and head it in parent.

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of

FODOR et al.

Continuation of Appln. No. 09/056,927 Group Art Unit: 1655

Filed: May 31, 2000 Examiner: S. Zitomer

FOR: PRODUCTS FOR DETECTING NUCLEIC ACIDS (as amended)

* * *

June 2, 2000

PRELIMINARY AMENDMENT

Honorable Commissioner of
Patents and Trademarks
Washington, D.C. 20231

Sir:

Please enter the following amendments before examination of this application.

IN THE TITLE:

Change the title, wherever it appears (e.g., cover page, page 1 and page 143), to read --PRODUCTS FOR DETECTING NUCLEIC ACIDS--.

IN THE SPECIFICATION:

Amend the specification as follows.

Page 1, after the title and before the first paragraph, insert --This is a continuation of Appln. No. 09/056,927, filed April 8, 1998, pending; which is a continuation of Appln. No. 08/670,118, filed June 25, 1996, now U.S. Pat. No. 5,800,992; which is a divisional of Appln. No. 08/168,904, filed December 15, 1993; which is a continuation of Appln. No. 07/624,114, filed December 6, 1990, now abandoned; which is a continuation in-part of commonly assigned Appln. No. 07/492,462, filed March 7, 1990, now U.S. Pat. No. 5,143,854; and Appln. No. 07/362,901, filed June 7, 1989, now abandoned which are hereby incorporated by reference.

Additional commonly assigned Appln. Nos. 07/624,120 and 07/626,730, both of which were filed on December 6, 1990; Appln. No. 07/435,316, filed November 13, 1989, now abandoned; and U.S. Pat. No. 5,252,743 are also hereby incorporated herein by

reference.--; and

Page 1, line 31, delete "the"; line 39, insert --of-- after "sequence".

Page 3, line 11, change "VLSIPS" to --VLSIPS™ Technology--; insert after "VLSIPS™ Technology" --(Very Large Scale Immobilized Polymer Synthesis)--.

Page 3 before the paragraph starting on line 16, insert the paragraph: --According to one aspect of the masking technique, the invention provides an ordered method for forming a plurality of polymer sequences by sequential addition of reagents comprising the step of serially protecting and deprotecting portions of the plurality of polymer sequences for addition of other portions of the polymer sequences using a binary synthesis strategy.--.

Page 6, line 13, delete "are".

Page 7, line 22, delete "upon".

Page 8, line 4, change "VLSIPS" to --VLSIPS™ Technology--; line 5, delete "peptide" and insert --nucleotide--; line 6, change "VLSIPS" to --VLSIPS™ Technology--; line 7, delete "dipeptide" and insert --nucleotide--; line 8, change "VLSIPS" to --VLSIPS™ Technology--.

Page 12, line 11, change "VLSIPS Substrates" to --VLSIPS™ Technology--; line 15, change "(VLSIPS)" to --(VLSIPS™)--; line 35, delete "U.S.S.N. 07/492,462 (VLSIPS CIP)" and insert --Pirrung et al. (1992) U.S. Pat. No. 5,143,854--.

Page 13, line 6, delete "U.S.S.N. 07/492,462 (VLSIPS CIP)" and insert --Pirrung et al. (1992) U.S. Pat. No. 5,143,854--; lines 6-7, insert in the blank spaces after "U.S.S.N." --07/624,120, now abandoned-- and delete "attorney docket number 11509-28 (automated VLSIPS)"; line 30, insert in the blank spaces --07/624,120, now abandoned--; delete "attorney docket number 11509-28 (automated VLSIPS)"; line 31, insert in the blank spaces after "U.S.S.N." --Dower et al. (1995) U.S. Pat. No. 5,427,908--; lines 31-32, delete "attorney docket number 11509-16 (sequencing by synthesis)".

Page 14, line 28, after "112:", delete "6997" and insert --6397--; insert --6399-- in the blank space after "6997".

Page 15, line 11, delete "U.S.S.N. 07/492,462 (VLSIPS CIP)" and insert --Pirrung et al. (1992) U.S. Pat. No. 5,143,854--; lines 14-15, insert in the blank spaces after "U.S. S.N." --07/624,120, now abandoned-- and delete "attorney docket number 11509-28 (automated VLSIPS)".

Page 20, line 12, after "07/362,901" insert --, now abandoned-- and delete " (VLSIPS parent)" and delete "U.S.S.N. 07/492,462" and insert --Pirrung et al. (1992) U.S. Pat. No.

5,143,854--; line 13, delete (VLSIPS CIP)" and also delete "(caged biotin parent)" and delete "Serial No. 07/612,671" and insert --Barrett et al. (1993) U.S. Pat. No. 5,252,743--; line 14, delete "(caged biotin CIP)"; line 36, after "07/362,901" insert --, now abandoned--; line 37, delete "U.S.S.N. 07/492,462" and insert --Pirrung et al. (1992) U.S. Pat. No. 5,143,854-- and delete "(VLSIPS parent)"; line 37, insert in the blank spaces after "U.S.S.N." --07/624,120, now abandoned--; line 37, delete "(VLSIPS CIP)"; lines 37-38, delete "attorney docket number 11509-28 (automated VLSIPS)".

Page 21, line 11, insert in the blank spaces after "U.S.S.N." --07/624,120, now abandoned--; lines 11-12, delete "attorney docket number 11509-28 (automated VLSIPS)".

Page 22, line 3, change "statistically" to --statistical--; line 24, change "Khropko" to --Khrapko--.

Page 23, line 12, change "VLSIPS" to --VLSIPS™ Technology--.

Page 24, line 24, insert --the-- after "that".

Page 26, line 30, change "way," to --way--.

Page 31, line 27, insert in the blank spaces after "U.S.S.N." --07/624,120, now abandoned--; lines 27-28, delete "attorney docket number 11509-28 (automated VLSIPS)".

Page 33, line 14, after "07/435,316" insert --, now abandoned-- and delete "(caged biotin parent)" and delete "Serial No. 07/612,671" and insert --Barrett et al. (1993) U.S. Pat. No. 5,252,743--; line 15, delete "(caged biotin CIP)".

Page 34, line 14, change "affixed" to --affixing--; insert --them-- before the phrase "to a definable position"; line 16, after "07/362,901" insert --, now abandoned--; lines 16-17, delete "(VLSIPS parent)"; line 17, delete "Serial No. 07/492,462" and insert --Pirrung et al. (1992) U.S. Pat. No. 5,143,854--; and delete "(VLSIPS CIP)".

Page 35, line 4, insert --through the-- after "particularly"; line 9, delete "U.S.S.N. 07/492,462 (VLSIPS CIP)" and insert --Pirrung et al. (1992) U.S. Pat. No. 5,143,854--; line 10, insert in the blank spaces after "U.S.S.N." --07/624,120, now abandoned--; lines 10-11, delete "attorney docket number 11509-28 (automated VLSIPS)"; line 11, delete "U.S.S.N. 07/612,671 (caged biotin CIP)" and insert --Barrett et al. (1993) U.S. Pat. No. 5,252,743--; line 16, delete "U.S.S.N. 07/492,462" and insert --Pirrung et al. (1992) U.S. Pat. No. 5,143,854--; line 17, insert in the blank spaces after "U.S.S.N." --07/624,120, now abandoned--; line 17, delete "(VLSIPS CIP)"; lines 17-18, delete "attorney docket number 11509-28 (automated VLSIPS)"; line 30, delete "U.S.S.N. 07/492,462 (VLSIPS CIP)" and

insert --Pirrung et al. (1992) U.S. Pat. No. 5,143,854--; line 31, insert in the blank spaces after "U.S.S.N." --07/624,120, now abandoned--; lines 31-32, delete "attorney docket number 11509-28 (automated VLSIPS)".

Page 36, line 9, change "are" to --is--; line 21, insert in the blank spaces --07/624,120, now abandoned--; delete "attorney docket number 11509-28 (automated VLSIPS)"; line 32, insert in the blank spaces after "U.S.S.N." --07/624,120, now abandoned--; line 33, delete "attorney docket number 11509-28 (automated VLSIPS)".

Page 37, line 12, change "converse" to --confer--; line 30, in the blank space after "6333-" insert --6335--; line 31, insert in the blank space --196--.

Page 38, line 15, insert --the-- before "target".

Page 39, line 14, change "it will" to --will allow the probe to--; line 16, change "close" to --closely--.

Page 42, line 21, insert --such-- before "that"; line 26, insert --such-- before "that"; line 31, change "VLSIPS" to --VLSIPS™ Technology--; line 35, change "VLSIPS" to --VLSIPS™ Technology--.

Page 41, line 22, delete "arylammonium" and insert --alkylammonium --; line 29, delete "arylammonium" and insert --alkylammonium--.

Page 42, line 37-38, delete "U.S.S.N. 07/492,462" and insert --Pirrung et al. (1992) U.S. Pat. No. 5,143,854--.

Page 43, line 9, insert in the blank spaces after "U.S.S.N." --07/624,120, now abandoned--; line 9, delete "(VLSIPS CIP)"; lines 9-10, delete "attorney docket number 11509-28 (automated VLSIPS)".

Page 47, line 38, delete "length of the".

Page 48, line 12, change "individual" to --individually--; line 38, insert in the blank spaces after "U.S.S.N." --07/626,730--.

Page 48, line 38 and page 49, line 1, delete "attorney docket number 11509-26 (sequencing by synthesis)".

Page 49, line 12, in the blank space after "341-" insert --347--.

Page 50, line 26, change "VLSIPS" to --VLSIPS™ Technology--.

Page 51, line 4, change "VLSIPS" to --VLSIPS™ Technology--; line 12, change "complex, it" to --complex. It--.

Page 52, line 13, delete the first occurrence of "expressed".

Page 55, line 3, insert --and-- after “susceptibilities,".

Page 56, line 7, change "fingerprinted" to --fingerprinting--; change "as in" to --an--; line 8, change "mosaism" to --mosaicism--.

Page 58, line 4, delete the sentence "Again, the target sequences may be desired to be fragmented"; insert the sentence --Again, it may be desirable to fragment the target sequences.--.

Page 60, line 19, change "VLSIPS" to --VLSIPS™ Technology--; line 33, delete the second occurrence of "the".

Page 62, line 3, insert --a-- after "provides"; line 10, change "to define" to --in defining--; line 20, change "as" to --are--.

Page 63, line 8, change "upon" to --by--; line 29, change "amendable" to --detectable--; line 38, change "present" to --presently available--.

Page 64, line 1, change "screen" to --screens--; line 17, change "also is" to --is also--; lines 33-34, delete "U.S.S.N. 07/492,462" and insert --Pirrung et al. (1992) U.S. Pat. No. 5,143,854--; line 34, insert in the blank spaces after "U.S.S.N." --07/624,120, now abandoned-- and delete "(VLSIPS CIP)"; lines 34-35, delete "attorney docket number 11509-28 (automated VLSIPS)".

Page 65, line 3, change "VLSIPS" to --VLSIPS™ Technology--; line 9, delete "U.S.S.N. 07/462,492 (VLSIPS CIP)" and insert --Pirrung et al. (1992) U.S. Pat. No. 5,143,854--; line 10, insert in the blank spaces --07/624,120, now abandoned--; delete "attorney docket number 11509-28 (automated VLSIPS)".

Page 66, line 36, insert in the blank spaces after "U.S.S.N." --07/624,120, now abandoned--; lines 36-37, delete "attorney docket number 11509-28 (automated VLSIPS)".

Page 67, line 14, change "then" to --them--; line 33, change "VLSIPS" to --VLSIPS™ Technology--.

Page 68, line 1, delete "U.S.S.N. 07/492,462 (VLSIPS CIP)" and insert --Pirrung et al. (1992) U.S. Pat. No. 5,143,854--; line 2, insert in the blank spaces --07/624,120, now abandoned-- and delete "attorney docket number 11509-28 (automated VLSIPS)"; line 7, insert in the blank spaces after "U.S.S.N." --07/624,120, now abandoned--; lines 7-8, delete "attorney docket number 11509-28 (automated VLSIPS)"; line 33, delete "U.S.S.N. 07/492,462 (automated VLSIPS)" and insert --Pirrung et al. (1992) U.S. Pat. No. 5,143,854--; line 34, insert in the blank spaces after "U.S.S.N." --07/624,120, now abandoned--; lines

34-35, delete "attorney docket number 11509-28 (automated VLSIPS)"; line 38, change "VLSIPS" to --VLSIPS™ Technology--.

Page 70, line 2, insert --,-- after "Fig. C"; line 6, change "anaines" to --amines--.

Page 71, line 24, change "have been" to --include, for example--.

Page 73, line 12, insert --,-- after "functionalized"; line 13, insert --:-- after "(see example below)"; line 33, change "group" to --groups--; line 34, insert --:-- after "acetate".

Page 74, line 3, change "." to --:--; line 16, change "Related to" to --With respect to--; line 30, change '.' to --:--.

Page 75, line 4, insert in the blank spaces after "U.S.S.N." --07/624,120, now abandoned--; line 5, delete "attorney docket number 11509-28"; line 21, after "07/435,316" insert --, now abandoned--; lines 21-22, delete "(caged biotin parent)"; line 22, delete "U.S.S.N. 07/612,671 (caged biotin CIP)" and insert --Barrett et al. (1993) U.S. Pat. No. 5,252,743--; line 32, change "VLSIPS" to --VLSIPS™ Technology--; line 34, change ""standard VLSIPS"" to --"standard VLSIPS™ Technology"--; lines 34-35, insert in the blank spaces after "U.S.S.N." --07/624,120, now abandoned-- and delete "attorney docket number 11509-28 (automated VLSIPS)".

Page 77, line 2, delete "in our labs"; line 7, change "Brosystems" to --Biosystems--.

Page 78, line 27, delete "U.S.S.N. 07/612,671 (caged biotin CIP)" and insert --Barrett et al. (1993) U.S. Pat. No. 5,252,743--.

Page 79, line 31, change "VLSIPS" to --VLSIPS™ Technology--.

Page 81, line 12, insert --be-- after "may"; line 33, change "Cell" to --In addition to cell--.

Page 82, line 11, change "VLSIPS" to --VLSIPS™--; line 19, change "VLSIPS" to --VLSIPS™ Technology--.

Page 84, line 27, delete ")".

Page 86, line 30, change "ficoerythrin" to --phycoerythrin--; line 32, change "primary" to --primarily--.

Page 89, line 16, delete "U.S.S.N. 07/649,642 (VLSIPS CIP)" and insert --Pirrung et al. (1992) U.S. Pat. No. 5,143,854--; line 17, insert in the blank spaces after "U.S.S.N." --07/624,120, now abandoned--; lines 17-18, delete "attorney docket number 11509-28 (automated VLSIPS)"; line 25, insert in the blank spaces after "U.S.S.N." --07/624,120, now abandoned--; lines 25-26, delete "attorney docket number 11509-28 (automated VLSIPS)".

Page 90, line 21, insert in the blank spaces after "U.S.S.N." --07/624,120, now abandoned--; lines 21-22, delete "attorney docket number 11509-28 (automated VLSIPS)"; line 36, insert --a-- between "of" and "vast'".

Page 91, line 7, delete "up".

Page 94, line 29, change "VLSIPS" to --VLSIPS™ Technology--; line 31, change "simply" to --simplify--.

Page 95, line 15, insert --to-- after "adjacent".

Page 96, line 20, insert --for-- after "sequentially".

Page 97, line 6, delete "of".

Page 98, line 26, change "VLSIPS" to --VLSIPS™--; lines 28-29, delete "Serial No. 07/612,671 (caged biotin CIP)" and insert --Barrett et al. (1993) U.S. Pat. No. 5,252,743--; line 34, change "aa" to --a--.

Page 100, line 12, change "An" to --In--.

Page 101, line 10, change "preventative" to --preventive--.

Page 103, line 38, delete "(VLSIPS parent)" and add --, now abandoned-- after "1989"; line 39, delete "U.S.S.N. 07/492,462 (VLSIPS CIP)" and insert --Pirrung et al. (1992) U.S. Pat. No. 5,143,854--; line 40, after "07/435,316" insert --, now abandoned--; and delete "(caged biotin)"; line 41, delete "Serial No. 07/612,671 (caged biotin CIP)" and insert --Barrett et al. (1993) U.S. Pat. No. 5,252,743--; line 43, insert in the blank spaces after "U.S.S.N." --07/624,120, now abandoned--; lines 43-44, delete "attorney docket number 11509-28 (automated VLSIPS)"; line 45, insert in the blank spaces --07/626,730--; lines 45-46, delete "attorney docket number 11509-26 (sequencing by synthesis)".

Page 104, line 34, change "0-nitrovertryl" to --0-nitroveratryl--; line 35, change "0-nitrovertryl" to --0-nitroveratryl--.

Page 105, line 31, change "VLSIPS" to --VLSIPS™ Technology--.

Page 107, line 5, insert --to-- after "added"; line 13, delete "a".

Page 111, line 34, delete "U.S.S.N. 07/492,462 (VLSIPS CIP)" and insert --Pirrung et al. (1992) U.S. Pat. No. 5,143,854--; line 35, insert in the blank spaces after "U.S.S.N." --07,624,120, now abandoned--; lines 35-36, delete "attorney docket number 11509-28 (automated VLSIPS)".

Page 112, line 23, delete "U.S.S.N. 07/492,462 (VLSIPS CIP)" and insert --Pirrung et al. (1992) U.S. Pat. No. 5,143,854--; and insert in the blank spaces after "U.S.S.N."

--07/624,120, now abandoned--; lines 23-24, delete "attorney docket number 11509-28 (automated VLSIPS)".

Page 114, lines 21-22, delete "U.S.S.N. 07/492,462 (VLSIPS CIP)" and insert --Pirrung et al. (1992) U.S. Pat. No. 5,143,854--; line 22, insert in the blank spaces after "U.S.S.N." --07/624,120, now abandoned--; lines 22-23, delete "attorney docket number 11509-28".

Page 115, lines 10-11, delete "U.S.S.N. 07/492,462 (VLSIPS CIP)" and insert --Pirrung et al. (1992) U.S. Pat. No. 5,143,854--; line 23, insert in the blank spaces after "U.S.S.N." --07/624,120, now abandoned--; lines 23-24, delete "attorney docket number 11509-28"; line 36, delete "U.S.S.N. 07/492,462 (VLSIPS CIP)" and insert --Pirrung et al. (1992) U.S. Pat. No. 5,143,854--.

Page 116, line 8, change "durapore" to --Durapore™--.

Page 117, lines 15-16, delete "U.S.S.N. 07/492,462 (VLSIPS CIP)" and insert --Pirrung et al. (1992) U.S. Pat. No. 5,143,854--;

Page 118, line 14, delete "U.S.S.N. 07/492,462 (VLSIPS CIP)" and insert --Pirrung et al. (1992) U.S. Pat. No. 5,143,854--; line 22, change "NOVC" to --NVOC--; line 29, delete "U.S.S.N. 07/492,462 (VLSIPS CIP)" and insert --Pirrung et al. (1992) U.S. Pat No. 5,143,854--.

Page 120, line 8, delete "that"; line 35, delete "U.S.S.N. 07/492,462 (VLSIPS CIP)" and insert --Pirrung et al. (1992) U.S. Pat. No. 5,143,854--.

Page 124, line 26, after "07/435,316" insert --, now abandoned--; line 27, delete "(caged biotin parent)"; lines 27-28, delete "U.S.S.N. 07/612,671 (caged biotin TRIP)" and insert --Barrett et al. (1993) U.S. Pat. No. 5,252,743--.

Page 126, line 2, change "are" to --is--.

Page 127, line 3, delete "U.S.S.N. 07/612,671 (caged biotin CIP)" and insert --Barrett et al. (1993) U.S. Pat. No. 5,252,743--.

Page 131, line 4, insert --and-- after "recognition".

Page 133, line 2, insert --to-- after "environment".

Page 134, line 7, change "her" to --their--.

Page 135, line 37, change "VLSIPS" to --VLSIPS™ Technology--.

Page 136, line 22, change "VLSIPS" to --VLSIPS™ Technology--.

Page 144, delete the paragraph starting on line 15 with "This application", and ending on line 30 with "herein by reference".

IN THE CLAIMS:

Cancel claim 1 without prejudice and add the following new claims.

--26. A collection of beads comprised of different beads;
wherein a plurality of the beads have at least one polymer of a specific sequence attached thereto; and
wherein a plurality of the beads having at least one attached polymer are coded by an encoding system; and, the encoding system indicates the specific sequence of the polymer attached to a single bead.

27. The collection of claim 26, wherein the polymer attached to a single bead is an oligonucleotide having a given length; and wherein the oligonucleotide attached to a single bead is selected from the group consisting of all possible oligonucleotide sequences having the same number of nucleotides.

28. The collection of claim 27, wherein at least about 20% of all possible oligonucleotide sequences having the same number of nucleotides are attached to a different single bead.

29. The collection of claim 27, wherein at least about 70% of all possible oligonucleotide sequences having the same number of nucleotides are attached to a different single bead.

30. The collection of claim 27, wherein the oligonucleotide sequences having the same number of nucleotides are at least 5 nucleotides long.

31. The collection of claim 27, wherein the oligonucleotide sequences having the same number of nucleotides are 8 nucleotides long.

32. The collection of claim 27, wherein the oligonucleotide sequences having the same number of nucleotides are 9 nucleotides long.

33. The collection of claim 27, wherein the oligonucleotide sequences having the same number of nucleotides are at least 10 nucleotides long.

34. The collection of claim 27, wherein at least 10,000 of all the possible oligonucleotide sequences having the same number of nucleotides are attached to a different single bead.

35. The collection of claim 27, wherein at least 100,000 of all the possible oligonucleotide sequences having the same number of nucleotides are attached to a different single bead.

36. The collection of claim 27, wherein at least 1,000,000 of all the possible oligonucleotide sequences having the same number of nucleotides are attached to a different single bead.

37. The collection of claim 26, wherein the polymer is selected from the group consisting of polynucleotides and polypeptides.

38. The collection of claim 26, wherein the polymer is a protein selected from the group consisting of enzyme binding sites and antibody binding sites.

39. The collection of claim 26, wherein a plurality of beads are comprised of a glass surface and amines of poly-aminopropyltriethoxysilane thereon, and polymers are attached through amines on the glass surface.

40. The collection of claim 26, wherein the a plurality of beads are comprised of a surface and hydroxyl groups of an acrylic acid polymer thereon, and polymers are attached through hydroxyl groups on the surface.

41. The collection of claim 26, wherein the polymers are oligodeoxyribonucleotides, a plurality of a beads are comprised of a surface and a coating of an organic hydrophilic layer terminating in hydroxyl groups, and phosphates of the oligodeoxyribonucleotides are immediately linked to the hydroxyl groups.

42. The collection of claim 26, wherein the encoding system is selected from the group consisting of a magnetic system, a shape encoding system, a color encoding system, and combinations thereof.

43. A system for determining the nucleotide sequence of a target comprising:

(a) the collection of beads according to claim 27, wherein the attached oligonucleotides are complementary to substantially all possible oligonucleotide target sequences of a given length;

(b) an apparatus that sorts single beads that have bound an oligonucleotide target from single beads that have not bound an oligonucleotide target; and

(c) an apparatus that decodes by the encoding system to indicate the oligonucleotide sequence attached to a single bead.

44. The system of claim 43, wherein at least one of the targets is labeled with at least one detectable marker.

45. The system of claim 44, wherein the detectable marker is selected from the group consisting of fluorescent labels, radioisotopes, chemiluminescent compounds, bioluminescent sources, labeled binding proteins, heavy metal atoms, spectroscopic markers, magnetic labels, linked enzymes, chromogens, dyes, and spin labels.

46. The collection of claim 26, wherein the polymer attached to a single bead is an oligonucleotide probe having a given length.

47. A system for fingerprinting comprising:

(a) the collection of beads according to claim 46, wherein the attached oligonucleotide probes of a specific sequence are complementary to oligonucleotide targets;

(b) an apparatus that sorts single beads that have bound an oligonucleotide target from beads that have not bound an oligonucleotide target; and

(c) an apparatus that decodes the encoding system, to indicate the oligonucleotide sequence attached to a single bead.

48. The system of claim 47, wherein at least one of the targets is labeled with at least one detectable marker.

49. The system of claim 47, wherein the detectable marker is selected from the group consisting of fluorescent labels, radioisotopes, chemiluminescent compounds, bioluminescent sources, labeled binding proteins, heavy metal atoms, spectroscopic markers, magnetic labels, linked enzymes, chromogens, dyes, and spin labels.

50. The system of claim 47, wherein the oligonucleotide probes and the targets are greater than 25 nucleotides, and different fluorescent labels are the detectable markers.

51. The collection of claim 26, wherein the polymer is selected from the group consisting of enzyme binding sites and antibody binding sites.

52. A system for fingerprinting, comprising:

(a) the collection of beads according to claim 26, wherein the polymer is a polypeptide able to specifically bind a target;

(b) an apparatus that sorts single beads that have bound a target from beads that have not bound a target; and

(c) an apparatus that decodes the encoding system to indicate the polypeptide sequence attached to a single bead.

53. The collection of claim 26, wherein a plurality of beads are comprised of a TEFLON copolymer graft surface and a coating of a hydrophilic organic layer terminating in hydroxyl sites.

54. The collection of claim 27, wherein a plurality of the attached oligonucleotides are comprised of at least one nucleotide analogue.

55. The collection of claim 27, wherein a plurality of the different beads are reusable; thereby allowing specific interactions between the polymer attached to a single bead and its target to be disrupted, the single bead treated, whereby a renewed plurality of beads equivalent to an unused plurality of beads is made by such treatment.

56. The collection of claim 27, wherein the given length of the oligonucleotide attached to a plurality of beads is selected from the group consisting of oligonucleotide sequences having a variable number of nucleotides.

57. A collection of fibers comprised of different fibers;
wherein a plurality of the fibers have at least one polymer of a specific sequence attached thereto; and
wherein a plurality of the fibers having at least one attached polymer are coded by an encoding system; and, the encoding system indicates the specific sequence of the polymer attached to a single fiber.

58. The collection of claim 57, wherein the polymer attached to a single fiber is an oligonucleotide having a given length; and wherein the oligonucleotide attached to a single fiber is selected from the group consisting of all possible oligonucleotide sequences having the same number of nucleotides.

59. The collection of claim 58, wherein at least about 25% of all possible oligonucleotide sequences having the same number of nucleotides are attached to a different single fiber.

60. The collection of claim 58, wherein at least about 70% of all possible oligonucleotide sequences having the same number of nucleotides are attached to a different single fiber.

61. The collection of claim 57 wherein the polymer is selected from the group consisting of polynucleotides and polypeptides.

62. The collection of claim 57, wherein the polymer is a protein selected from the group consisting of enzyme binding sites and antibody binding sites.

63. A system for determining the nucleotide sequence of a target comprising:
(a) the collection of fibers according to claim 57, wherein the attached oligonucleotides are complementary to substantially all possible oligonucleotide target sequences of a given length;
(b) an apparatus that sorts single fibers that have bound an oligonucleotide target from single beads that have not bound an oligonucleotide target; and
(c) an apparatus that decodes the encoding system to indicate the oligonucleotide sequence attached to a single fiber.

64. The collection of the fibers of claim 57, wherein the polymer attached to a single fiber is an oligonucleotide probe having a given length.

65. A system for fingerprinting, comprising:
(a) the collection of fibers according to claim 64, wherein the attached oligonucleotide probes of a specific sequence are complementary to oligonucleotide targets;
(b) an apparatus that sorts single fibers that have bound an oligonucleotide target from fibers that have not bound an oligonucleotide target; and
(c) an apparatus that decodes the encoding system to indicate the oligonucleotide sequence attached to a single fiber.

66. The collection of claim 57, wherein a plurality of fibers are comprised of a TEFLON copolymer graft surface and a coating of a hydrophilic organic layer terminating in hydroxyl sites.

67. The collection of fibers of claim 58, wherein a plurality of the attached oligonucleotides are comprised of at least one nucleotide analogue.

68. The collection of claim 57, wherein a plurality of the different fibers are reusable; thereby allowing specific interactions between the polymer attached to a single fiber and its target to be disrupted, the single fiber treated, whereby a renewed plurality of fibers equivalent to an unused plurality of fibers is made by such treatment.

69. The collection of claim 58, wherein the given length of the oligonucleotide attached to a plurality of fibers is selected from the group consisting of oligonucleotide sequences having a variable number of nucleotides.--

REMARKS

Claims 26-69 are pending. These product claims are useful for the practice of the methods claimed in co-pending Appln. No. 09/362,089.

The amendments to the specification and claims find support throughout the original disclosure. The new claims are directed to an invention disclosed in the specification, but not originally claimed. It is submitted that no new matter has been added by these amendments.

In particular, the Examiner's attention is directed to the attached table entitled "Support for New Claims" showing support in the original specification for the new claims.

Many of the above amendments to the specification were entered as a preliminary amendment in parent U.S. Appln. No. 08/670,118. Given the number of amendments made to the specification, the Examiner in that case required a substitute specification. Therefore, to expedite prosecution of the present application, Applicants will be submitting a substitute specification with the amendments made in this Preliminary Amendment after an Official filing receipt has been received. At that time, new formal drawings will also be submitted.

An early and favorable examination on the merits is earnestly requested.

Respectfully submitted,

PILLSBURY MADISON & SUTRO, L.L.P.

By _____ 43,180
for Paul N. Kokulis
Reg. No. 16,773
Telephone: (202) 861-3503
Facsimile: (202) 822-0944

Enclosure

Claims	Original Specification Support (Col:Line)
26. A collection of beads:comprised of different beads; wherein a plurality of the beads have at least one polymer of a specific sequence attached thereto; and wherein a plurality of the beads having at least one attached polymer are coded by an encoding system; and, the encoding system indicates the specific sequence of the polymer attached to a single bead.	The present invention provides a composition comprising a plurality of positionally distinguishable sequence specific reagents attached to a solid substrate, which reagents are capable of specifically binding to a predetermined subunit sequence of a preselected multisubunit length.... 3:21-32 The invention provides methods for sequencing a polymer . . . . In one embodiment, the substrates are beads. 5:1-15 It should be noted that multiple substrates may be simultaneously exposed to a single target sequence where each substrate is a duplicate of one another where, in combination, multiple substrates together provide the complete or desired subset of possible subsequences. 38:1-27 [E]ach probe might be attached to a single bead or substrate and the beads sorted by whether there is a binding interaction. Those beads which do bind might be encoded to indicate the subsequence specificity of reagents attached thereto. . . . [A] sorting system may be utilized . . . . 38:1-27 The . . . method utilizes synthetic beads or fibers. 76:1 – 77:10
27. The collection of claim 26, wherein the polymer attached to a single bead is an oligonucleotide having a given length; and wherein the oligonucleotide attached to a single bead is selected from the group consisting of all possible oligonucleotide sequences having the same number of nucleotides.	[T]he plurality of reagents comprise substantially all possible subsequences of said preselected length found in said target. 5:16-18 In a bead embodiment, at least some of the plurality of substrates have one subsequence specific reagent attached thereto, and the substrates are coded to indicate the sequence specificity of said reagent. 5:24-28 The enablement of the sequencing process by hybridization is based in large part upon the ability to synthesize a large number (e.g., to virtually saturate) of the possible overlapping sequence segments.... 7:8-11 Once the desired repertoire of possible oligomer sequences of a given length have been synthesized, this collection of reagents may be individually positionally attached to a substrate, thereby allowing a batchwise hybridization step. 34:34-38
28. The collection of claim 27, wherein at least about 20% of all possible oligonucleotide sequences having the same number of nucleotides are attached to a different single bead.	About 20% would be preferred.... 26:23-24

29. The collection of claim 27, wherein at least about 70% of all possible oligonucleotide sequences having the same number of nucleotides are attached to a different single bead.	In particular, although a substantial fraction will preferably be at least about 70%.... 26:19-25
30. The collection of claim 27, wherein the oligonucleotide sequences having the same number of nucleotides are at least 5 nucleotides long.	The number of possible five digit subsequences is 2^5 = 32. The number of possible different sequences 10 digits long is 2^{10} = 1,024. The five contiguous digit subsequences within a ten digit sequence number six, i.e., positioned at digits 1-5, 2-6, 3-7, 4-8, 5-9, and 6-10. 23:15-20 [I]n an absolute sequencing application, it is often useful to have virtually all of the possible oligonucleotides of a given length. 38:33-34
31. The collection of claim 27, wherein the oligonucleotide sequences having the same number of nucleotides are 8 nucleotides long.	The number of possible five digit subsequences is 2^5 = 32. The number of possible different sequences 10 digits long is 2^{10} = 1,024. The five contiguous digit subsequences within a ten digit sequence number six, i.e., positioned at digits 1-5, 2-6, 3-7, 4-8, 5-9, and 6-10. 23:15-20 As indicated above, there are 65,536 8-mers, 262,144 9-mers, 1,048,576 10-mers.... 38:35-36
32. The collection of claim 27, wherein the oligonucleotide sequences having the same number of nucleotides are 9 nucleotides long.	As indicated above, there are 65,536 8-mers, 262,144 9-mers, 1,048,576 10-mers.... 38:35-36
33. The collection of claim 27, wherein the oligonucleotide sequences having the same number of nucleotides are at least 10 nucleotides long.	As indicated above, there are 65,536 8-mers, 262,144 9-mers, 1,048,576 10-mers.... 38:35-36
34. The collection of claim 27, wherein at least 10,000 of all the possible oligonucleotide sequences having the same number of nucleotides are attached to a different single bead.	[O]ligonucleotide probes ... including numbers in excess of about 10^2, 10^3, 10^4, 10^5, 10^6, or even more.... 13:15-17
35. The collection of claim 27, wherein at least 100,000 of all the possible oligonucleotide sequences having the same number of nucleotides are attached to a different single bead.	[O]ligonucleotide probes ... including numbers in excess of about 10^2, 10^3, 10^4, 10^5, 10^6, or even more.... 13:15-17
36. The collection of claim 27, wherein at least 1,000,000 of all the possible oligonucleotide sequences having the same number of nucleotides are attached to a different single bead.	[O]ligonucleotide probes ... including numbers in excess of about 10^2, 10^3, 10^4, 10^5, 10^6, or even more.... 13:15-17
37. The collection of claim 26, wherein the polymer is selected from the group consisting of polynucleotides and polypeptides.	The present invention provides a composition comprising a plurality of positionally distinguishable sequence specific reagents attached to a solid subtrate.... In some embodiments, the subunit sequence is a polynucleotide or a polypeptide. 3:21-28 The specific sequence recognition reagents will typically

	be oligonucleotide probes which hybridize with specificity to subsequences found on the target sequence. 11:12-15 These reagents will take the form, typically, of proteins exhibiting binding specificity.... 44:30-33
38. The collection of claim 26, wherein the polymer is a protein selected from the group consisting of enzyme binding sites and antibody binding sites.	[T]he nonpolynucleotide sequences typically require other sequence recognition reagents. These reagents will take the form, typically, of proteins exhibiting binding specificity, e.g., enzyme binding site or antibody binding sites. 44:28-33
39. The collection of claim 26, wherein a plurality of beads are comprised of a glass surface and amines of poly-aminopropyltriethoxysilane thereon, and polymers are attached through amines on the glass surface.	The ... "standard VLSIPS" method... involves applying amino-propyltriethoxysilane to a glass surface. 75: 34-37 The polymeric substrate approach involves either of two ways of generating a polymeric substrate. The first uses a high concentration of aminopropyltriethoxysilane.... This... provides a high density of amines on the surface of the glass. 76:1-7
40. The collection of claim 26, wherein the a plurality of beads are comprised of a surface and hydroxyl groups of an acrylic acid polymer thereon, and polymers are attached through hydroxyl groups on the surface.	The second method involves either the coating or covalent binding of an appropriate acrylic acid polymers onto the substrate surface. In particular... a monomer... is used to generate a high density of hydroxyl groups on the substrate surface, allowing for the formation of phosphate bonds.... Here the building up of, e.g., a DNA oligomer, can be started immediately since phosphate bonds to the surface can be accomplished in the first step with no need for modification. 76:11-37
41. The collection of claim 26, wherein the polymers are oligodeoxyribonucleotides, a plurality of a beads are comprised of a surface and a coating of an organic hydrophilic layer terminating in hydroxyl groups, and phosphates of the oligodeoxyribonucleotides are immediately linked to the hydroxyl groups.	The fourth method uses beads or fibers. This would use another substrate, such as teflon copolymer graft bead or fiber, which is covalently coated with an organic layer (hydrophilic) terminating in hydroxyl sites... allowing for immediate phosphate linkages.... 77:3-7
42. The collection of claim 26, wherein the encoding system is selected from the group consisting of a magnetic system, a shape encoding system, a color encoding system, and combinations thereof.	An encoding system may include a magnetic system, a shape encoding system, a color encoding system, or a combination of these, or any other encoding system. 38:22-24
43. A system for determining the nucleotide sequence of a target comprising: (a) the collection of beads according to claim 27, wherein the attached oligonucleotides are complementary to substantially all possible oligonucleotide target sequences of a given length; (b) an apparatus that sorts single beads that have bound an oligonucleotide target from single beads that have not bound an oligonucleotide target; and (c) an apparatus that decodes by the encoding system to indicate the oligonucleotide sequence attached to a single bead.	Thus, a less expensive, highly reliable, and labor efficient means for sequencing biological macromolecules is needed.... In particular, an automated system would improve the reproducibility and accuracy of procedures. The present invention satisfies these and other needs.... The present invention provides improved methods useful for de novo sequencing of an unknown polymer sequence, for verification of known sequences, for fingerprinting polymers, and for mapping homologous segments within a sequence. 2:26-3:2. It should be noted that multiple substrates may be simultaneously exposed to a single target sequence where each substrate is a duplicate of one another where, in combination, multiple substrates together provide the

	complete or desired subset of possible subsequences. This provides the opportunity to overcome a limitation of the density of positions on a single substrate by using multiple substrates. 38:1-8 In the extreme case, each probe might be attached to a single bead or substrate and the beads sorted by whether there is a binding interaction. Those beads which do bind might be encoded to indicate the subsequence specificity of reagents attached thereto. Then the target may be bound to the whole collection of beads and those beads that have appropriate specific reagents on them will bind to target. 38:8-15 Then a sorting system may be utilized to sort those beads that actually bind the target from those that do not. This may be accomplished by presently available cell sorting devices or a similar apparatus.... [T]he encoding scheme may be read off to determine the specificity of the reagent on the bead. 38:15-27
44. The system of claim 43, wherein at least one of the targets is labeled with at least one detectable marker.	The label used to detect the target sequences will be determined, in part, by the detection methods being applied.... 40:1-5 The target polynucleotide may be labeled by any of a number of convenient detectable markers. A fluorescent label is preferred.... 85: 36-38
45. The system of claim 44, wherein the detectable marker is selected from the group consisting of fluorescent labels, radioisotopes, chemiluminescent compounds, bioluminescent sources, labeled binding proteins, heavy metal atoms, spectroscopic markers, magnetic labels, linked enzymes, chromogens, dyes, and spin labels.	Other potential labeling moieties include, radioisotopes, chemiluminescent compounds, labeled binding proteins, heavy metal atoms, spectroscopic markers, magnetic labels, and linked enzymes.... An intercalative dye... may be used.... Suitable chromogens will include molecules and compounds which absorb light in a distinctive range...or emit light.... 86:2-28 Alternatively, luciferins may be used be used in conjunction with luciferase or lucigenins to provide bioluminescence. Spin labels are provided by reporter molecules with an unpaired electron spin.... 88:34-38 Exemplary spin labels include organic free radicals, transitional metal complexes, particularly vanadium, copper, iron, and manganese, and the like. Exemplary spin labels include nitroxide free radicals. 89:1-4
46. The collection of claim 26, wherein the polymer attached to a single bead is an oligonucleotide probe having a given length.	The specific sequence recognition reagents will typically be oligonucleotide probes which hybridize with specificity to subsequences found on the target sequence. 11:12-15 The length of oligonucleotides used in sequencing applications will be selected on criteria.... 37:1-3

47. A system for fingerprinting, comprising: (a) the collection of beads according to claim 46, wherein the attached oligonucleotide probes of a specific sequence are complementary to oligonucleotide targets; (b) an apparatus that sorts single beads that have bound an oligonucleotide target from beads that have not bound an oligonucleotide target; and (c) an apparatus that decodes the encoding system, to indicate the oligonucleotide sequence attached to a single bead.	The present invention provides improved methods useful for de novo sequencing of an unknown polymer sequence, for verification of known sequences, for fingerprinting polymers, and for mapping homologous segments within a sequence. 2:26-3:2 Preferably, the plurality of reagents comprise substantially all possible subsequences of said preselected length found in said target.... In a bead embodiment, at least some of the plurality of substrates have one subsequence specific reagent attached thereto, and the substrates are coded to indicate the sequence specificity of said reagent. 5:14-28 Once the desired repertoire of possible oligomer sequences of a given length have been synthesized, this collection of reagents may be individually positionally attached to a substrate, thereby allowing a batchwise hybridization step. 34:34-38 It should be noted that multiple substrates may be simultaneously exposed to a single target sequence where each substrate is a duplicate of one another where, in combination, multiple substrates together provide the complete or desired subset of possible subsequences. This provides the opportunity to overcome a limitation of the density of positions on a single substrate by using multiple substrates. 38:1-8 In the extreme case, each probe might be attached to a single bead or substrate and the beads sorted by whether there is a binding interaction. Those beads which do bind might be encoded to indicate the subsequence specificity of reagents attached thereto. Then the target may be bound to the whole collection of beads and those beads that have appropriate specific reagents on them will bind to target. 38:8-15 Then a sorting system may be utilized to sort those beads that actually bind the target from those that do not. This may be accomplished by presently available cell sorting devices or a similar apparatus.... [T]he encoding scheme may be read off to determine the specificity of the reagent on the bead. 38:15-27 The hybridization conditions between probe and target should be selected such that the specific recognition interaction.... 41:20-30 These reagents will take the form, typically, of proteins.... 44:30-33 [A]t least four different substrate preparation procedures... synthetic beads or fibers. 75:30-33 The fourth method uses beads or fibers. 77:3-10

48. The system of claim 47, wherein at least one of the targets is labeled with at least one detectable marker.	The target polynucleotide may be labeled by any of a number of convenient detectable markers. A fluorescent label is preferred.... 85: 36-38 The label used to detect the target sequences will be determined, in part, by the detection methods being applied.... 40:1-5
49. The system of claim 47, wherein the detectable marker is selected from the group consisting of fluorescent labels, radioisotopes, chemiluminescent compounds, bioluminescent sources, labeled binding proteins, heavy metal atoms, spectroscopic markers, magnetic labels, linked enzymes, chromogens, dyes, and spin labels.	Other potential labeling moieties include, radioisotopes, chemiluminescent compounds, labeled binding proteins, heavy metal atoms, spectroscopic markers, magnetic labels, and linked enzymes.... An intercalative dye... may be used.... Suitable chromogens will include molecules and compounds which absorb light in a distinctive range...or emit light.... 86:2-28 Alternatively, luciferins may be used be used in conjunction with luciferase or lucigenins to provide bioluminescence. Spin labels are provided by reporter molecules with an unpaired electron spin.... 88:34-38 Exemplary spin labels include organic free radicals, transitional metal complexes, particularly vanadium, copper, iron, and manganese, and the like. Exemplary spin labels include nitroxide free radicals. 89:1-4
50. The system of claim 47, wherein the oligonucleotide probes and the targets are greater than 25 nucleotides, and different fluorescent labels are the detectable markers.	The target polynucleotide may be labeled by any of a number of convenient detectable markers. A fluorescent label is preferred.... 85: 36-38 [I]f oligonucleotide probes are being used, their lengths should be apporximately comparable and will be selected to hybridize.... [T]he target and oligonucleotide probes are of lengths typically greater than about 25 nucleotides. 50:8-13 In another embodiment, different targets may be simultaneously sequenced where each target has a different label. 86:18-20
51. The collection of claim 26, wherein the polymer is selected from the group consisting of enzyme binding sites and antibody binding sites.	[T]he nonpolynucleotide sequences typically require other sequence recognition reagents. These reagents will take the form, typically, of proteins exhibiting binding specificity, e.g., enzyme binding site or antibody binding sites. 44:28-33
52. A system for fingerprinting, comprising: (a) the collection of beads according to claim 26, wherein the polymer is a polypeptide able to specifically bind a target; (b) an apparatus that sorts single beads that have bound a target from beads that have not bound a target; and (c) an apparatus that decodes the encoding system to indicate the polypeptide sequence attached to a single bead.	The present invention provides improved methods useful for de novo sequencing of an unknown polymer sequence, for verification of known sequences, for fingerprinting polymers, and for mapping homologous segments within a sequence. 2:26-3:2 Preferably, the plurality of reagents comprise substantially all possible subsequences of said preselected length found in said target.... In a bead embodiment, at least some of the plurality of substrates have one subsequence specific reagent attached thereto, and the substrates are coded to indicate the sequence

	specificity of said reagent. 5:14-28 Once the desired repertoire of possible oligomer sequences of a given length have been synthesized, this collection of reagents may be individually positionally attached to a substrate, thereby allowing a batchwise hybridization step. 34:34-38 It should be noted that multiple substrates may be simultaneously exposed to a single target sequence where each substrate is a duplicate of one another where, in combination, multiple substrates together provide the complete or desired subset of possible subsequences. This provides the opportunity to overcome a limitation of the density of positions on a single substrate by using multiple substrates. 38:1-8 In the extreme case, each probe might be attached to a single bead or substrate and the beads sorted by whether there is a binding interaction. Those beads which do bind might be encoded to indicate the subsequence specificity of reagents attached thereto. Then the target may be bound to the whole collection of beads and those beads that have appropriate specific reagents on them will bind to target. 38:8-15 Then a sorting system may be utilized to sort those beads that actually bind the target from those that do not. This may be accomplished by presently available cell sorting devices or a similar apparatus.... [T]he encoding scheme may be read off to determine the specificity of the reagent on the bead. 38:15-27 The hybridization conditions between probe and target should be selected such that the specific recognition interaction.... 41:20-30 These reagents will take the form, typically, of proteins.... 44:30-33 [A]t least four different substrate preparation procedures... synthetic beads or fibers. 75:30-33 The fourth method uses beads or fibers. 77:3-10
53. The collection of claim 26, wherein a plurality of beads are comprised of a TEFLON copolymer graft surface and a coating of a hydrophilic organic layer terminating in hydroxyl sites.	This would use another substrate, such as a teflon copolymer graft bead or fiber, which is covalently coated with an organic layer (hydrophilic) terminating in hydroxyl sites.... 77: 3-7
54. The collection of claim 27, wherein a plurality of the attached oligonucleotides are comprised of at least one nucleotide analogue.	By use of non-natural targeting reagents, e.g., unusual nucleotide analogues which pair with other natural nucleotide interactions.... 48:26-33

55. The collection of claim 27, wherein a plurality of the different beads are reusable; thereby allowing specific interactions between the polymer attached to a single bead and its target to be disrupted, the single bead treated, whereby a renewed plurality of beads equivalent to an unused plurality of beads is made by such treatment.	As with the sequencing application, the fingerprinting usages may also take advantage of the reusability of the substrate. In this way, the interactions can be disrupted, the substrate treated, and the renewed substrate is equivalent to an unused substrate. 51:29-33 Where a substrate is made with specific reagents that are relatively insensitive to the handling and processing steps involved in a single cycle of use, the substrate may often be reused. 93:9-12
56. The collection of claim 27, wherein the given length of the oligonucleotide attached to a plurality of beads is selected from the group consisting of oligonucleotide sequences having a variable number of nucleotides.	It will be recognized that the probe oligonucleotides will preferably, but need not necessarily, be of identical length.... [B]ut may contain a plurality of probes of a known sequence. 26:13-19 [F]ingerprint probes of various lengths, or with specific defined non-matches may be used. 39:30-31
57. A collection of fibers comprised of different fibers; wherein a plurality of the fibers have at least one polymer of a specific sequence attached thereto; and wherein a plurality of the fibers having at least one attached polymer are coded by an encoding system; and, the encoding system indicates the specific sequence of the polymer attached to a single fiber.	The present invention provides improved methods useful for de novo sequencing of an unknown polymer sequence, for verification of known sequences, for fingerprinting polymers, and for mapping homologous segments within a sequence. 2:26-3:2 The present invention provides a composition comprising a plurality of positionally distinguishable sequence specific reagents attached to a solid substrate, which reagents are capable of specifically binding to a predetermined subunit sequence of a preselected multisubunit length.... 3:21-32 The invention provides methods for sequencing a polymer . . . . In one embodiment, the substrates are beads. 5:1-15 It should be noted that multiple substrates may be simultaneously exposed to a single target sequence where each substrate is a duplicate of one another where, in combination, multiple substrates together provide the complete or desired subset of possible subsequences. 38:1-27 [E]ach probe might be attached to a single bead or substrate and the beads sorted by whether there is a binding interaction. Those beads which do bind might be encoded to indicate the subsequence specificity of reagents attached thereto. . . . [A] sorting system may be utilized . . . . 38:1-27 [A]t least four different substrate preparation procedures... synthetic beads or fibers. 75:30-33 The fourth method uses beads or fibers. 77:3-10

58. The collection of claim 57, wherein the polymer attached to a single fiber is an oligonucleotide having a given length; and, wherein, the oligonucleotide attached to a single fiber is selected from the group consisting of all possible oligonucleotide sequences having the same number of nucleotides.	[T]he plurality of reagents comprise substantially all possible subsequences of said preselected length found in said target. 5:16-18 In a bead embodiment, at least some of the plurality of substrates have one subsequence specific reagent attached thereto, and the substrates are coded to indicate the sequence specificity of said reagent. 5:24-28 The enablement of the sequencing process by hybridization is based in large part upon the ability to synthesize a large number (e.g., to virtually saturate) of the possible overlapping sequence segments.... 7:8-11 Once the desired repertoire of possible oligomer sequences of a given length have been synthesized, this collection of reagents may be individually positionally attached to a substrate, thereby allowing a batchwise hybridization step. 34:34-38
59. The collection of claim 58, wherein at least about 25% of all possible oligonucleotide sequences having the same number of nucleotides are attached to a different single fiber.	In other embodiments, the reagents represent[ ] at least about 25% of the possible subsequences of said preselected length. 3:36-38
60. The collection of claim 58, wherein at least about 70% of all possible oligonucleotide sequences having the same number of nucleotides are attached to a different single fiber.	In particular, although a substantial fraction will preferably be at least about 70%.... 26:19-25
61. The collection of claim 57 wherein the polymer is selected from the group consisting of polynucleotides and polypeptides.	The present invention provides a composition comprising a plurality of positionally distinguishable sequence specific reagents attached to a solid subtrate.... In some embodiments, the subunit sequence is a polynucleotide or a polypeptide. 3:21-28 The specific sequence recognition reagents will typically be oligonucleotide probes which hybridize with specificity to subsequences found on the target sequence. 11:12-15 These reagents will take the form, typically, of proteins exhibiting binding specificity.... 44:30-33
62. The collection of claim 57, wherein the polymer is a protein selected from the group consisting of enzyme binding sites and antibody binding sites.	[T]he nonpolynucleotide sequences typically require other sequence recognition reagents. These reagents will take the form, typically, of proteins exhibiting binding specificity, e.g., enzyme binding site or antibody binding sites. 44:28-33
63. A system for determining the nucleotide sequence of a target comprising: (a) the collection of fibers according to claim 57, wherein the attached oligonucleotides are complementary to substantially all possible oligonucleotide target sequences of a given length;	The present invention provides improved methods useful for de novo sequencing of an unknown polymer sequence, for verification of known sequences, for fingerprinting polymers, and for mapping homologous segments within a sequence. 2:26-3:2

(b) an apparatus that sorts single fibers that have bound an oligonucleotide target from single beads that have not bound an oligonucleotide target; and (c) an apparatus that decodes the encoding system to indicate the oligonucleotide sequence attached to a single fiber.	Preferably, the plurality of reagents comprise substantially all possible subsequences of said preselected length found in said target.... In a bead embodiment, at least some of the plurality of substrates have one subsequence specific reagent attached thereto, and the substrates are coded to indicate the sequence specificity of said reagent. 5:14-28 Once the desired repertoire of possible oligomer sequences of a given length have been synthesized, this collection of reagents may be individually positionally attached to a substrate, thereby allowing a batchwise hybridization step. 34:34-38 It should be noted that multiple substrates may be simultaneously exposed to a single target sequence where each substrate is a duplicate of one another where, in combination, multiple substrates together provide the complete or desired subset of possible subsequences. This provides the opportunity to overcome a limitation of the density of positions on a single substrate by using multiple substrates. 38:1-8 In the extreme case, each probe might be attached to a single bead or substrate and the beads sorted by whether there is a binding interaction. Those beads which do bind might be encoded to indicate the subsequence specificity of reagents attached thereto. Then the target may be bound to the whole collection of beads and those beads that have appropriate specific reagents on them will bind to target. 38:8-15 Then a sorting system may be utilized to sort those beads that actually bind the target from those that do not. This may be accomplished by presently available cell sorting devices or a similar apparatus.... [T]he encoding scheme may be read off to determine the specificity of the reagent on the bead. 38:15-27 The hybridization conditions between probe and target should be selected such that the specific recognition interaction.... 41:20-30 These reagents will take the form, typically, of proteins.... 44:30-33
64. The collection of the fibers of claim 57, wherein the polymer attached to a single fiber is an oligonucleotide probe having a given length.	[R]eagents will typically be oligonucleotide probes which hybridize with specificity to subsequences found on the target sequence. 11:11-15 The length of the oligonucleotide used in sequencing applications will be selected on criteria determines to some extent by the practical limits discussed above. 37:1-3

65. A system for fingerprinting, comprising: (a) the collection of fibers according to claim 64, wherein the attached oligonucleotide probes of a specific sequence are complementary to oligonucleotide targets; (b) an apparatus that sorts single fibers that have bound an oligonucleotide target from fibers that have not bound an oligonucleotide target; and (c) an apparatus that decodes the encoding system to indicate the oligonucleotide sequence attached to a single fiber.	The present invention provides improved methods useful for de novo sequencing of an unknown polymer sequence, for verification of known sequences, for fingerprinting polymers, and for mapping homologous segments within a sequence. 2:26-3:2 Preferably, the plurality of reagents comprise substantially all possible subsequences of said preselected length found in said target.... In a bead embodiment, at least some of the plurality of substrates have one subsequence specific reagent attached thereto, and the substrates are coded to indicate the sequence specificity of said reagent. 5:14-28 Once the desired repertoire of possible oligomer sequences of a given length have been synthesized, this collection of reagents may be individually positionally attached to a substrate, thereby allowing a batchwise hybridization step. 34:34-38 It should be noted that multiple substrates may be simultaneously exposed to a single target sequence where each substrate is a duplicate of one another where, in combination, multiple substrates together provide the complete or desired subset of possible subsequences. This provides the opportunity to overcome a limitation of the density of positions on a single substrate by using multiple substrates. 38:1-8 In the extreme case, each probe might be attached to a single bead or substrate and the beads sorted by whether there is a binding interaction. Those beads which do bind might be encoded to indicate the subsequence specificity of reagents attached thereto. Then the target may be bound to the whole collection of beads and those beads that have appropriate specific reagents on them will bind to target. 38:8-15 Then a sorting system may be utilized to sort those beads that actually bind the target from those that do not. This may be accomplished by presently available cell sorting devices or a similar apparatus.... [T]he encoding scheme may be read off to determine the specificity of the reagent on the bead. 38:15-27 The hybridization conditions between probe and target should be selected such that the specific recognition interaction.... 41:20-30 These reagents will take the form, typically, of proteins.... 44:30-33
66. The collection of claim 57, wherein a plurality of fibers are comprised of a TEFLON copolymer graft surface and a coating of a hydrophilic organic layer terminating in hydroxyl sites.	This would use another substrate, such as a teflon copolymer graft bead or fiber, which is covalently coated with an organic layer (hydrophilic) terminating in hydroxyl sites.... 77: 3-7

67. The collection of fibers of claim 58, wherein a plurality of the attached oligonucleotides are comprised of at least one nucleotide analogue.	By use of non-natural targeting reagents, e.g., unusual nucleotide analogues which pair with other natural nucleotide interactions.... 48:26-33
68. The collection of claim 57, wherein a plurality of the different fibers are reusable; thereby allowing specific interactions between the polymer attached to a single fiber and its target to be disrupted, the single fiber treated, whereby a renewed plurality of fibers equivalent to an unused plurality of fibers is made by such treatment.	As with the sequencing application, the fingerprinting usages may also take advantage of the reusability of the substrate. In this way, the interactions can be disrupted, the substrate treated, and the renewed substrate is equivalent to an unused substrate. 51:29-33 Where a substrate is made with specific reagents that are relatively insensitive to the handling and processing steps involved in a single cycle of use, the substrate may often be reused. 93:9-12
69. The collection of claim 58, wherein the given length of the oligonucleotide attached to a plurality of fibers is selected from the group consisting of oligonucleotide sequences having a variable number of nucleotides.	It will be recognized that the probe oligonucleotides will preferably, but need not necessarily, be of identical length.... [B]ut may contain a plurality of probes of a known sequence. 26:13-19 [F]ingerprint probes of various lengths, or with specific defined non-matches may be used. 39:30-31

Attorney Docket No. 11509-30

PATENT APPLICATION

SEQUENCING BY HYBRIDIZATION OF A TARGET NUCLEIC ACID TO A MATRIX OF DEFINED OLIGONUCLEOTIDES

Inventors:

Stephen P.A. Fodor, a citizen of the United States, and a resident of Palo Alto, California;

Dennis Solas, a citizen of the United States, and a resident of San Francisco, California, and

William J. Dower, a citizen of the United States, and a resident of Menlo Park, California.

Assignee:

Affymax Technologies, N.V.

TOWNSEND and TOWNSEND
Steuart Street Tower, 20th Floor
One Market Plaza
San Francisco, California 94105
(415) 543-9600

PATENT

Attorney Docket No. 11509-30

SEQUENCING BY HYBRIDIZATION OF A TARGET NUCLEIC ACID TO A MATRIX OF DEFINED OLIGONUCLEOTIDES

BACKGROUND OF THE INVENTION

The present invention relates to the sequencing, fingerprinting, and mapping of polymers, particularly biological polymers. The inventions may be applied, for example, in the sequencing, fingerprinting, or mapping of nucleic acids, polypeptides, oligosaccharides, and synthetic polymers.

The relationship between structure and function of macromolecules is of fundamental importance in the understanding of biological systems. These relationships are important to understanding, for example, the functions of enzymes, structural proteins, and signalling proteins, ways in which cells communicate with each other, as well as mechanisms of cellular control and metabolic feedback.

Genetic information is critical in continuation of life processes. Life is substantially informationally based and its genetic content controls the growth and reproduction of the organism and its complements. Polypeptides, which are critical features of all living systems, are encoded by the genetic material of the cell. In particular, the properties of enzymes, functional proteins, and structural proteins are determined by the sequence of amino acids which make them up. As structure and function are integrally related, many biological functions may be explained by elucidating the underlying the structural features which provide those functions. For this reason, it has become very important to determine the genetic sequences of nucleotides which encode the enzymes, structural proteins, and other effectors of biological functions. In addition to segments of nucleotides which encode polypeptides, there are many nucleotide sequences which are involved in control and regulation of gene expression.

The human genome project is directed toward determining the complete sequence the genome of the human

organism. Although such a sequence would not correspond to the sequence of any specific individual, it would provide significant information as to the general organization and specific sequences contained within segments from particular individuals. It would also provide mapping information which is very useful for further detailed studies. However, the need for highly rapid, accurate, and inexpensive sequencing technology is nowhere more apparent than in a demanding sequencing project such as this. To complete the sequencing of a human genome would require the determination of approximately $3x10^9$, or 3 billion base pairs.

The procedures typically used today for sequencing include the Sanger dideoxy method, see, e.g., Sanger et al. (1977) Proc. Natl. Acad. Sci. USA, 74:5463-5467, or the Maxam and Gilbert method, see, e.g., Maxam et al., (1980) Methods in Enzymology, 65:499-559. The Sanger method utilizes enzymatic elongation procedures with chain terminating nucleotides. The Maxam and Gilbert method uses chemical reactions exhibiting specificity of reaction to generate nucleotide specific cleavages. Both methods require a practitioner to perform a large number of complex manual manipulations. These manipulations usually require isolating homogeneous DNA fragments, elaborate and tedious preparing of samples, preparing a separating gel, applying samples to the gel, electrophoresing the samples into this gel, working up the finished gel, and analyzing the results of the procedure.

Thus, a less expensive, highly reliable, and labor efficient means for sequencing biological macromolecules is needed. A substantial reduction in cost and increase in speed of nucleotide sequencing would be very much welcomed. In particular, an automated system would improve the reproducibility and accuracy of procedures. The present invention satisfies these and other needs.

SUMMARY OF THE INVENTION

The present invention provides improved methods useful for de novo sequencing of an unknown polymer sequence, for verification of known sequences, for fingerprinting

polymers, and for mapping homologous segments within a sequence. By reducing the number of manual manipulations required and automating most of the steps, the speed, accuracy, and reliability of these procedures are greatly enhanced.

The production of a substrate having a matrix of positionally defined regions with attached reagents exhibiting known recognition specificity can be used for the sequence analysis of a polymer. Although most directly applicable to sequencing, the present invention is also applicable to fingerprinting, mapping, and general screening of specific interactions. The VLSIPS substrates will be applied to evaluating other polymers, e.g., carbohydrates, polypeptides, hydrocarbon synthetic polymers, and the like. For these non-polynucleotides, the sequence specific reagents will usually be antibodies specific for a particular subunit sequence.

The present invention also provides a means to automate sequencing manipulations. The automation of the substrate production method and of the scan and analysis steps minimizes the need for human intervention. This simplifies the tasks and promotes reproducibility.

The present invention provides a composition comprising a plurality of positionally distinguishable sequence specific reagents attached to a solid substrate, which reagents are capable of specifically binding to a predetermined subunit sequence of a preselected multi-subunit length having at least three subunits, said reagents representing substantially all possible sequences of said preselected length. In some embodiments, the subunit sequence is a polynucleotide or a polypeptide, in others the preselected multi-subunit length is five subunits and the subunit sequence is a polynucleotide sequence. In other embodiments, the specific reagent is an oligonucleotide of at least about five nucleotides. Alternatively, the specific reagent is a monoclonal antibody. Usually the specific reagents are all attached to a single solid substrate, and the reagents comprise about 3000 different sequences. In other embodiments, the reagents represents at least about 25% of the possible subsequences of said preselected length. Usually, the reagents are localized in

regions of the substrate having a density of at least 25 regions per square centimeter, and often the substrate has a surface area of less than about 4 square centimeters.

The present invention also provides methods for analyzing a sequence of a polynucleotide or a polypeptide, said method comprising the step of:

a) exposing said polynucleotide or polypeptide to a composition as described.

It also provides useful methods for identifying or comparing a target sequence with a reference, said method comprising the step of:

a) exposing said target sequence to a composition as described;

b) determining the pattern of positions of the reagents which specifically interact with the target sequence; and

c) comparing the pattern with the pattern exhibited by the reference when exposed to the composition.

The present invention also provides methods for sequencing a segment of a polynucleotide comprising the steps of:

a) combining:

i) a substrate comprising a plurality of chemically synthesized and positionally distinguishable oligonucleotides capable of recognizing defined oligonucleotide sequences; and

ii) a target polynucleotide; thereby forming high fidelity matched duplex structures of complementary subsequences of known sequence; and

b) determining which of said reagents have specifically interacted with subsequences in said target polynucleotide.

In one embodiment, the segment is substantially the entire length of said polynucleotide.

The invention also provides methods for sequencing a polymer, said method comprising the steps of:

a) preparing a plurality of reagents which each specifically bind to a subsequence of preselected length;

b) positionally attaching each of said reagents to one or more solid phase substrates, thereby producing substrates of positionally definable sequence specific probes;

c) combining said substrates with a target polymer whose sequence is to be determined; and

d) determining which of said reagents have specifically interacted with subsequences in said target polymer.

In one embodiment, the substrates are beads. Preferably, the plurality of reagents comprise substantially all possible subsequences of said preselected length found in said target. In another embodiment, the solid phase substrate is a single substrate having attached thereto reagents recognizing substantially all possible subsequences of preselected length found in said target.

In another embodiment, the method further comprises the step of analyzing a plurality of said recognized subsequences to assemble a sequence of said target polymer. In a bead embodiment, at least some of the plurality of substrates have one subsequence specific reagent attached thereto, and the substrates are coded to indicate the sequence specificity of said reagent.

The present invention also embraces a method of using a fluorescent nucleotide to detect interactions with oligonucleotide probes of known sequence, said method comprising:

a) attaching said nucleotide to a target unknown polynucleotide sequence, and

b) exposing said target polynucleotide sequence to a collection of positionally defined oligonucleotide probes of known sequences to

determine the sequences of said probes which interact with said target.

In a further refinement, an additional step is included of:

a) collating said known sequences to determine the overlaps of said known sequences to determine the sequence of said target sequence.

A method of mapping a plurality of sequences relative to one another is also provided, the method comprising:

a) preparing a substrate having a plurality of positionally attached sequence specific probes are attached;

b) exposing each of said sequences to said substrate, thereby determining the patterns of interaction between said sequence specific probes and said sequences; and

c) determining the relative locations of said sequence specific probe interactions on said sequences to determine the overlaps and order of said sequences.

In one refinement, the sequence specific probes are oligonucleotides, applicable to where the target sequences are nucleic acid sequences.

In the nucleic acid sequencing application, the steps of the sequencing process comprise:

a) producing a matrix substrate having known positionally defined regions of known sequence specific oligonucleotide probes;

b) hybridizing a target polynucleotide to the positions on the matrix so that each of the positions which contain oligonucleotide probes complementary to a sequence on the target hybridize to the target molecule;

c) detecting which positions have bound the target, thereby determining sequences which are found on the target; and

d) analyzing the known sequences contained in the target to determine sequence overlaps and assembling the sequence of the target therefrom.

The enablement of the sequencing process by hybridization is based in large part upon the ability to synthesize a large number (e.g., to virtually saturate) of the possible overlapping sequence segments and distinguishing those probes which hybridize with fidelity from those which have mismatched bases, and to analyze a highly complex pattern of hybridization results to determine the overlap regions.

The detecting of the positions which bind the target sequence would typically be through a fluorescent label on the target. Although a fluorescent label is probably most convenient, other sorts of labels, e.g., radioactive, enzyme linked, optically detectable, or spectroscopic labels may be used. Because the oligonucleotide probes are positionally defined, the location of the hybridized duplex will directly translate to the sequences which hybridize. Thus, upon analysis of the positions provides a collection of subsequences found within the target sequence. These subsequences are matched with respect to their overlaps so as to assemble an intact target sequence.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 illustrates a flow chart for sequence, fingerprint, or mapping analysis.

Fig. 2 illustrates the proper function of a VLSIPS peptide synthesis.

Fig. 3 illustrates the proper function of a VLSIPS dipeptide synthesis.

Fig. 4 illustrates the process of a VLSIPS trinucleotide synthesis.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

I. Overall Description
- A. general
- B. VLSIPS substrates
- C. binary masking
- D. applications
- E. detection methods and apparatus
- F. data analysis

II. Theoretical Analysis
- A. simple n-mer structure; theory
- B. complications
- C. non-polynucleotide embodiments

III. Polynucleotide Sequencing
- A. preparation of substrate matrix
- B. labeling target polynucleotide
- C. hybridization conditions
- D. detection; VLSIPS scanning
- E. analysis
- F. substrate reuse
- G. non-polynucleotide aspects

IV. Fingerprinting
- A. general
- B. preparation of substrate matrix
- C. labeling target nucleotides
- D. hybridization conditions
- E. detection; VLSIPS scanning
- F. analysis
- G. substrate reuse
- H. non-polynucleotide aspects

V. Mapping
- A. general
- B. preparation of substrate matrix
- C. labeling
- D. hybridization/specific interaction
- E. detection
- F. analysis
- G. substrate reuse
- H. non-polynucleotide aspects

VI. Additional Screening
- A. specific interactions
- B. sequence comparisons
- C. categorizations
- D. statistical correlations

VII. Formation of Substrate
- A. instrumentation
- B. binary masking
- C. synthetic methods
- D. surface immobilization

* * *

I. OVERALL DESCRIPTION

A. General

The present invention relies in part on the ability to synthesize or attach specific recognition reagents at known locations on a substrate, typically a single substrate. In particular, the present invention provides the ability to prepare a substrate having a very high density matrix pattern of positionally defined specific recognition reagents. The reagents are capable of interacting with their specific targets while attached to the substrate, e.g., solid phase interactions, and by appropriate labeling of these targets, the sites of the interactions between the target and the specific reagents may be derived. Because the reagents are positionally defined, the sites of the interactions will define the specificity of each interaction. As a result, a map of the patterns of interactions with specific reagents on the substrate is convertible into information on the specific interactions taking place, e.g., the recognized features.

Where the specific reagents recognize a large number of possible features, this system allows the determination of the combination of specific interactions which exist on the target molecule. Where the number of features is sufficiently large, the identical same combination, or pattern, of features is sufficiently unlikely that a particular target molecule may often be uniquely defined by its features. In the extreme, the features may actually be the subunit sequence of the target molecule, and a given target sequence may be uniquely defined by its combination of features.

In particular, the methodology is applicable to sequencing polynucleotides. The specific sequence recognition reagents will typically be oligonucleotide probes which hybridize with specificity to subsequences found on the target sequence. A sufficiently large number of those probes allows the fingerprinting of a target polynucleotide or the relative mapping of a collection of target polynucleotides, as described in greater detail below.

In the high resolution fingerprinting provided by a saturating collection of probes which include all possible subsequences of a given size, e.g., 10-mers, collating of all the subsequences and determination of specific overlaps will be derived and the entire sequence can usually be reconstructed.

Although a polynucleotide sequence analysis is a preferred embodiment, for which the specific reagents are most easily accessible, the invention is also applicable to analysis of other polymers, including polypeptides, carbohydrates, and synthetic polymers, including α-, β-, and ω-amino acids, polyurethanes, polyesters, polycarbonates, polyureas, polyamides, polyethyleneimines, polyarylene sulfides, polysiloxanes, polyimides, polyacetates, and mixed polymers. Various optical isomers, e.g., various D- and L- forms of the monomers, may be used.

Sequence analysis will take the form of complete sequence determination, to the level of the sequence of individual subunits along the entire length of the target sequence. Sequence analysis also takes the form of sequence homology, e.g., less than absolute subunit resolution, where

"similarity" in the sequence will be detectable, or the form of selective sequences of homology interspersed at specific or irregular locations.

In either case, the sequence is determinable at selective resolution or at particular locations. Thus, the hybridization method will be useful as a means for identification, e.g., a "fingerprint", much like a Southern hybridization method is used. It is also useful to map particular target sequences.

B. VLSIPS Substrates

The invention is enabled by the development of technology to prepare substrates on which specific reagents may be either positionally attached or synthesized. In particular, the very large scale immobilized polymer synthesis (VLSIPS) technology allows for the very high density production of an enormous diversity of reagents mapped out in a known matrix pattern on a substrate. These reagents specifically recognize subsequences in a target polymer and bind thereto, producing a map of positionally defined regions of interaction. These map positions are convertible into actual features recognized, and thus would be present in the target molecule of interest.

As indicated, the sequence specific recognition reagents will often be oligonucleotides which hybridize with fidelity and discrimination to the target sequence. For use with other polymers, monoclonal or polyclonal antibodies having high sequence specificity will often be used.

In the generic sense, the VLSIPS technology allows the production of a substrate with a high density matrix of positionally mapped regions with specific recognition reagents attached at each distinct region. By use of protective groups which can be positionally removed, or added, the regions can be activated or deactivated for addition of particular reagents or compounds. Details of the protection are described below and in related application U.S.S.N. 07/492,462 (VLSIPS CIP). In a preferred embodiment, photosensitive protecting agents will be used and the regions of activation or deactivation may be controlled by electro-optical and optical methods, similar to

many of the processes used in semiconductor wafer and chip fabrication.

In the nucleic acid nucleotide sequencing application, a VLSIPS substrate is synthesized having positionally defined oligonucleotide probes. See U.S.S.N. 07/492,462 (VLSIPS CIP); and U.S.S.N. __/___,___, attorney docket number 11509-28 (automated VLSIPS). By use of masking technology and photosensitive synthetic subunits, the VLSIPS apparatus allows for the stepwise synthesis of polymers according to a positionally defined matrix pattern. Each oligonucleotide probe will be synthesized at known and defined positional locations on the substrate. This forms a matrix pattern of known relationship between position and specificity of interaction. The VLSIPS technology allows the production of a very large number of different oligonucleotide probes to be simultaneously and automatically synthesized including numbers in excess of about 10^2, 10^3, 10^4, 10^5, 10^6, or even more, and at densities of at least about 10^2, $10^3/cm^2$, $10^4/cm^2$, $10^5/cm^2$ and up to $10^6/cm^2$ or more. This application discloses methods for synthesizing polymers on a silicon or other suitably derivatized substrate, methods and chemistry for synthesizing specific types of biological polymers on those substrates, apparatus for scanning and detecting whether interaction has occurred at specific locations on the substrate, and various other technologies related to the use of a high density very large scale immobilized polymer substrate. In particular, sequencing, fingerprinting, and mapping applications are discussed herein in detail, though related technologies are described in simultaneously filed applications U.S.S.N. __/___,___, attorney docket number 11509-28 (automated VLSIPS) and U.S.S.N. __/___,___, attorney docket number 11509-16 (sequencing by synthesis), each of which is hereby incorporated herein by reference.

In other embodiments, antibody probes will be generated which specifically recognize particular subsequences found on a polymer. Antibodies would be generated which are specific for recognizing a three contiguous amino acid sequence, and monoclonal antibodies may be preferred.

Optimally, these antibodies would not recognize any sequences other than the specific three amino acid stretch desired and the binding affinity should be insensitive to flanking or remote sequences found on a target molecule. Likewise, antibodies specific for particular carbohydrate linkages or sequences will be generated. A similar approach could be used for preparing specific reagents which recognize other polymer subunit sequences. These reagents would typically be site specifically localized to a substrate matrix pattern where the regions are closely packed.

These reagents could be individually attached at specific sites on the substrate in a matrix by an automated procedure where the regions are positionally targeted by some other specific mechanism, e.g., one which would allow the entire collection of reagents to be attached to the substrate in a single reaction. Each reagent could be separately attached to a specific oligonucleotide sequence by an automated procedure. This would produce a collection of reagents where, e.g., each monoclonal antibody would have a unique oligonucleotide sequence attached to it. By virtue of a VLSIPS substrate which has different complementary oligonucleotides synthesized on it, each monoclonal antibody would specifically be bound only at that site on the substrate where the complementary oligonucleotide has been synthesized. A crosslinking step would fix the reagent to the substrate. See, e.g., Dattagupta et al. (1985) U.S. Patent No. 4,542,102 and (1987) U.S. Pat. No. 4,713,326; and Chatterjee, M. et al. (1990) J. Am. Chem. Soc. 112:6997-____, which are hereby incorporated herein by reference. This allows a high density positionally specific collection of specific recognition reagents, e.g., monoclonal antibodies, to be immobilized to a solid substrate using an automated system.

The regions which define particular reagents will usually be generated by selective protecting groups which may be activated or deactivated. Typically the protecting group will be bound to a monomer subunit or spatial region, and can be spatially affected by an activator, such as electromagnetic radiation. Examples of protective groups with utility herein

include nitroveratryl oxycarbonyl (NVOC), nitrobenzyl oxycarbony (NBOC), dimethyl dimethoxy benzyloxy carbonyl, 5-bromo-7-nitroindolinyl, O-hydroxy-α-methyl cinnamoyl, and 2-oxymethylene anthraquinone. Examples of activators include ion beams, electric fields, magnetic fields, electron beams, x-ray, and other forms of electromagnetic radiation.

C. Binary Masking

In fact, the means for producing a substrate useful for these techniques are explained in U.S.S.N. 07/492,462 (VLSIPS CIP), which is hereby incorporated herein by reference. However, there are various particular ways to optimize the synthetic processes. Many of these methods are described in U.S.S.N. __/___,___, attorney docket number 11509-28 (automated VLSIPS).

Briefly, the binary synthesis strategy refers to an ordered strategy for parallel synthesis of diverse polymer sequences by sequential addition of reagents which may be represented by a reactant matrix, and a switch matrix, the product of which is a product matrix. A reactant matrix is a 1 x n matrix of the building blocks to be added. The switch matrix is all or a subset of the binary numbers from 1 to n arranged in columns. In preferred embodiments, a binary strategy is one in which at least two successive steps illuminate half of a region of interest on the substrate. In most preferred embodiments, binary synthesis refers to a synthesis strategy which also factors a previous addition step. For example, a strategy in which a switch matrix for a masking strategy halves regions that were previously illuminated, illuminating about half of the previously illuminated region and protecting the remaining half (while also protecting about half of previously protected regions and illuminating about half of previously protected regions). It will be recognized that binary rounds may be interspersed with non-binary rounds and that only a portion of a substrate may be subjected to a binary scheme, but will still be considered to be a binary masking scheme within the definition herein. A binary "masking" strategy is a binary synthesis which uses light to

remove protective groups from materials for addition of other materials such as nucleotides or amino acids.

In particular, this procedure provides a simplified and highly efficient method for saturating all possible sequences of a defined length polymer. This masking strategy is also particularly useful in producing all possible oligonucleotide sequence probes of a given length.

D. Applications

The technology provided by the present invention has very broad applications. Although described specifically for polynucleotide sequences, similar sequencing, fingerprinting, mapping, and screening procedures can be applied to polypeptide, carbohydrate, or other polymers. In particular, the present invention may be used to completely sequence a given target sequence to subunit resolution. This may be for de novo sequencing, or may be used in conjunction with a second sequencing procedure to provide independent verification. See, e.g., (1988) Science 242:1245. For example, a large polynucleotide sequence defined by either the Maxam and Gilbert technique or by the Sanger technique may be verified by using the present invention.

In addition, by selection of appropriate probes, a polynucleotide sequence can be fingerprinted. Fingerprinting is a less detailed sequence analysis which usually involves the characterization of a sequence by a combination of defined features. Sequence fingerprinting is particularly useful because the repertoire of possible features which can be tested is virtually infinite. Moreover, the stringency of matching is also variable depending upon the application. A Southern Blot analysis may be characterized as a means of simple fingerprint analysis.

Fingerprinting analysis may be performed to the resolution of specific nucleotides, or may be used to determine homologies, most commonly for large segments. In particular, an array of oligonucleotide probes of virtually any workable size may be positionally localized on a matrix and used to probe a sequence for either absolute complementary matching, or

homology to the desired level of stringency using selected hybridization conditions.

In addition, the present invention provides means for mapping analysis of a target sequence or sequences. Mapping will usually involve the sequential ordering of a plurality of various sequences, or may involve the localization of a particular sequence within a plurality of sequences. This may be achieved by immobilizing particular large segments onto the matrix and probing with a shorter sequence to determine which of the large sequences contain that smaller sequence. Alternatively, relatively shorter probes of known or random sequence may be immobilized to the matrix and a map of various different target sequences may be determined from overlaps. Principles of such an approach are described in some detail by Evans et al. (1989) "Physical Mapping of Complex Genomes by Cosmid Multiplex Analysis," Proc. Natl. Acad. Sci. USA 86:5030-5034; Michiels et al. (1987) "Molecular Approaches to Genome Analysis: A Strategy for the Construction of Ordered Overlap Clone Libraries," CABIOS 3:203-210; Olsen et al. (1986) "Random-Clone Strategy for Genomic Restriction Mapping in Yeast," Proc. Natl. Acad. Sci. USA 83:7826-7830; Craig, et al. (1990) "Ordering of Cosmid Clones Covering the Herpes Simplex Virus Type I (HSV-I) Genome: A Test Case for Fingerprinting by Hybridization," Nuc. Acids Res. 18:2653-2660; and Coulson, et al. (1986) "Toward a Physical Map of the Genome of the Nematode Caenorhabditis elegans," Proc. Natl. Acad. Sci. USA 83:7821-7825; each of which is hereby incorporated herein by reference.

Fingerprinting analysis also provides a means of identification. In addition to its value in apprehension of criminals from whom a biological sample, e.g., blood, has been collected, fingerprinting can ensure personal identification for other reasons. For example, it may be useful for identification of bodies in tragedies such as fire, flood, and vehicle crashes. In other cases the identification may be useful in identification of persons suffering from amnesia, or of missing persons. Other forensics applications include establishing the identity of a person, e.g., military identification "dog tags", or may be used in identifying the

source of particular biological samples. Fingerprinting technology is described, e.g., in Carrano, et al. (1989) "A High-Resolution, Fluorescence-Based, Semi-automated method for DNA Fingerprinting," Genomics 4: 129-136, which is hereby incorporated herein by reference. See, e.g., table I, for nucleic acid applications, and corresponding applications may be accomplished using polypeptides.

TABLE I

VLSIPS PROJECT IN NUCLEIC ACIDS

I. Construction of Chips

II. Applications

- A. Sequencing
  - 1. Primary sequencing
  - 2. Secondary sequencing (sequence checking)
  - 3. Large scale mapping
  - 4. Fingerprinting
- B. Duplex/Triplex formation
  - 1. Antisense
  - 2. Sequence specific function modulation (e.g. promoter inhibition)
- C. Diagnosis
  - 1. Genetic markers
  - 2. Type markers
    - a. Blood donors
    - b. Tissue transplants
- D. Microbiology
  - 1. Clinical microbiology
  - 2. Food microbiology

III. Instrumentation

- A. Chip machines
- B. Detection

IV. Software Development

- A. Instrumentation software
- B. Data reduction software
- C. Sequence analysis software

The fingerprinting analysis may be used to perform various types of genetic screening. For example, a single substrate may be generated with a plurality of screening probes, allowing for the simultaneous genetic screening for a large number of genetic markers. Thus, prenatal or diagnostic screening can be simplified, economized, and made more generally accessible.

In addition to the sequencing, fingerprinting, and mapping applications, the present invention also provides means for determining specificity of interaction with particular sequences. Many of these applications were described in U.S.S.N. 07/362,901 (VLSIPS parent), U.S.S.N. 07/492,462 (VLSIPS CIP), U.S.S.N. 07/435,316 (caged biotin parent), and U.S.S.N. 07/612,671 (caged biotin CIP).

E. Detection Methods and Apparatus

An appropriate detection method applicable to the selected labeling method can be selected. Suitable labels include radionucleotides, enzymes, substrates, cofactors, inhibitors, magnetic particles, heavy metal atoms, and particularly fluorescers, chemiluminescers, and spectroscopic labels. Patents teaching the use of such labels include U.S. Patent Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149; and 4,366,241.

With an appropriate label selected, the detection system best adapted for high resolution and high sensitivity detection may be selected. As indicated above, an optically detectable system, e.g., fluorescence or chemiluminescence would be preferred. Other detection systems may be adapted to the purpose, e.g., electron microscopy, scanning electron microscopy (SEM), scanning tunneling electron microscopy (STEM), infrared microscopy, atomic force microscopy (AFM), electrical condutance, and image plate transfer.

With a detection method selected, an apparatus for scanning the substrate will be designed. Apparatus, as described in U.S.S.N. 07/362,901 (VLSIPS parent); or U.S.S.N. 07/492,462 (VLSIPS CIP); or U.S.S.N. __/___,___, attorney docket number 11509-28 (automated VLSIPS), are particularly

appropriate. Design modifications may also be incorporated therein.

F. Data Analysis

Data is analyzed by processes similar to those described below in the section describing theoretical analysis. More efficient algorithms will be mathematically devised, and will usually be designed to be performed on a computer. Various computer programs which may more quickly or efficiently make measurement samples and distinguish signal from noise will also be devised. See, particularly, U.S.S.N. __/___,___, attorney docket number 11509-28 (automated VLSIPS).

The initial data resulting from the detection system is an array of data indicative of fluorescent intensity versus location on the substrate. The data are typically taken over regions substantially smaller than the area in which synthesis of a given polymer has taken place. Merely by way of example, if polymers were synthesized in squares on the substrate having dimensions of 500 microns by 500 microns, the data may be taken over regions having dimensions of 5 microns by 5 microns. In most preferred embodiments, the regions over which florescence data are taken across the substrate are less than about 1/2 the area of the regions in which individual polymers are synthesized, preferably less than 1/10 the area in which a single polymer is synthesized, and most preferably less than 1/100 the area in which a single polymer is synthesized. Hence, within any area in which a given polymer has been synthesized, a large number of fluorescence data points are collected.

A plot of number of pixels versus intensity for a scan should bear a rough resemblance to a bell curve, but spurious data are observed, particularly at higher intensities. Since it is desirable to use an average of fluorescent intensity over a given synthesis region in determining relative binding affinity, these spurious data will tend to undesirably skew the data.

Accordingly, in one embodiment of the invention the data are corrected for removal of these spurious data points,

and an average of the data points is thereafter utilized in determining relative binding efficiency. In general the data are fitted to a base curve and statistically measures are used to remove spurious data.

In an additional analytical tool, various degeneracy reducing analogues may be incorporated in the hybridization probes. Various aspects of this strategy are described, e.g., in Macevicz, S. (1990) PCT publication number WO 90/04652, which is hereby incorporated herein by reference.

II. THEORETICAL ANALYSIS

The principle of the hybridization sequencing procedure is based, in part, upon the ability to determine overlaps of short segments. The VLSIPS technology provides the ability to generate reagents which will saturate the possible short subsequence recognition possibilities. The principle is most easily illustrated by using a binary sequence, such as a sequence of zeros and ones. Once having illustrated the application to a binary alphabet, the principle may easily be understood to encompass three letter, four letter, five or more letter, even 20 letter alphabets. A theoretical treatment of analysis of subsequence information to reconstruction of a target sequence is provided, e.e., in Lysov, Yu., et al. (1988) Doklady Akademi. Nauk. SSR 303:1508-1511; Khropko K., et al. (1989) FEBS Letters 256:118-122; Pevzner, P. (1989) J. of Biomolecular Structure and Dynamics 7:63-69; and Drmanac, R. et al. (1989) Genomics 4:114-128; each of which is hereby incorporated herein by reference.

The reagents for recognizing the subsequences will usually be specific for recognizing a particular polymer subsequence anywhere within a target polymer. It is preferable that conditions may be devised which allow absolute discrimination between high fidelity matching and very low levels of mismatching. The reagent interaction will preferably exhibit no sensitivity to flanking sequences, to the subsequence position within the target, or to any other remote structure within the sequence. For polynucleotide sequencing, the specific reagents can be oligonucleotide probes; for

polypeptides and carbohydrates, antibodies will be useful reagents. Antibody reagents should also be useful for other types of polymers.

A. Simple n-mer Structure: Theory

1. Simple two letter alphabet: example

A simple example is presented below of how a sequence of ten digits comprising zeros and ones would be sequenceable using short segments of five digits. For example, consider the sample ten digit sequence:

1010011100.

A VLSIPS substrate could be constructed, as discussed elsewhere, which would have reagents attached in a defined matrix pattern which specifically recognize each of the possible five digit sequences of ones and zeros. The number of possible five digit subsequences is $2^5 = 32$. The number of possible different sequences 10 digits long is $2^{10} = 1,024$. The five contiguous digit subsequences within a ten digit sequence number six, i.e., positioned at digits 1-5, 2-6, 3-7, 4-8, 5-9, and 6-10. It will be noted that the specific order of the digits in the sequence is important and that the order is directional, e.g., running left to right versus right to left. The first five digit sequence contained in the target sequence is 10100. The second is 01001, the third is 10011, the fourth is 00111, the fifth is 01110, and the sixth is 11100.

The VLSIPS substrate would have a matrix pattern of positionally attached reagents which recognize each of the different 5-mer subsequences. Those reagents which recognize each of the 6 contained 5-mers will bind the target, and a label allows the positional determination of where the sequence specific interaction has occurred. By correlation of the position in the matrix pattern, the corresponding bound subsequences can be determined.

In the above-mentioned sequence, six different 5-mer sequences would be determined to be present. They would be:

10100
01001
10011
00111
01110
11100

Any sequence which contains the first five digit sequence, 10100, already narrows the number of possible sequences (e.g., from 1024 possible sequences) which contain it to less than about 192 possible sequences.

This 192 is derived from the observation that with the subsequence 10100 at the far left of the sequence, in positions 1-5, there are only 32 possible sequences. Likewise, for that particular subsequence in positions 2-6, 3-7, 4-8, 5-9, and 6-10. So, to sum up all of the sequences that could contain 10100, there are 32 for each position and 6 positions for a total of about 192 possible sequences. However, some of these 10 digit sequences will have been counted twice. Thus, by virtue of containing the 10100 subsequence, the number of possible 10-mer sequences has been decreased from 1024 sequences to less than about 192 sequences.

In this example, not only do we know that sequence contains 10100, but we also know that it contains the second five character sequence, 01001. By virtue of knowing that the sequence contains 10100, we can look specifically to determine whether the sequence contains a subsequence of five characters which contains the four leftmost digits plus a next digit to the left. For example, we would look for a sequence of X1010, but we find that there is none. Thus, we know that the 10100 must be at the left end of the 10-mer. We would also look to see whether the sequence contains the rightmost four digits plus a next digit to the right, e.g., 0100X. We find that the sequence also contains the sequence 01001, and that X is a 1. Thus, we know at least that our target sequence has an overlap of 0100 and has the left terminal sequence 101001.

Applying the same procedure to the second 5-mer, we also know that the sequence must include a sequence of five digits having the sequence 1001Y where Y must be either 0 or 1.

We look through the fragments and we see that we have a 10011 sequence within our target, thus Y is also 1. Thus, we would know that our sequence has a sequence of the first seven being 1010011.

Moving to the next 5-mer, we know that there must be a sequence of 0011Z, where Z must be either 0 or 1. We look at the fragments produced above and see that the target sequence contains a 00111 subsequence and Z is 1. Thus, we know the sequence must start with 10100111.

The next 5-mer must be of the sequence 0111W where W must be 0 or 1. Again, looking up at the fragments produced, we see that the target sequence contains a 01110 subsequence, and W is a 0. Thus, our sequence to this point is 101001110. We know that the last 5-mer must be either 11100 or 11101. Looking above, we see that it is 11100 and that must be the last of our sequence. Thus, we have determined that our sequence must have been 1010011100.

However, it will be recognized from the example above with the sequences provided therein, that the sequence analysis can start with any known positive probe subsequence. The determination may be performed by moving linearly along the sequence checking the known sequence with a limited number of next positions. Given this possibility, the sequence may be determined, besides by scanning all possible oligonucleotide probe positions, by specifically looking only where the next possible positions would be. This may increase the complexity of the scanning but may provide a longer time span dedicated towards scanning and detecting specific positions of interest relative to other sequence possibilities. Thus, the scanning apparatus could be set up to work its way along a sequence from a given contained oligonucleotide to only look at those positions on the substrate which are expected to have a positive signal.

It is seen that given a sequence, it can be de-constructed into n-mers to produce a set of internal contiguous subsequences. From any given target sequence, we would be able to determine what fragments would result. The hybridization sequence method depends, in part, upon being able to work in

the reverse, from a set of fragments of known sequences to the full sequence. In simple cases, one is able to start at a single position and work in either or both directions towards the ends of the sequence as illustrated in the example.

The number of possible sequences of a given length increases very quickly with the length of that sequence. Thus, a 10-mer of zeros and ones has 1024 possibilities, a 12-mer has 4096. A 20-mer has over a million possibilities, and a 30-mer has over a billion. However, a given 30-mer has, at most, 26 different internal 5-mer sequences. Thus, a 30 character target sequence having over a million possible sequences can be substantially defined by only 26 different 5-mers. It will be recognized that the probe oligonucleotides will preferably, but need not necessarily, be of identical length, and that the probe sequences need not necessarily be contiguous in that the overlapping subsequences need not differ by only a single subunit. Moreover, each position of the matrix pattern need not be homogeneous, but may actually contain a plurality of probes of known sequence. In addition, although all of the possible subsequence specifications would be preferred, a less than full set of sequences specifications could be used. In particular, although a substantial fraction will preferably be at least about 70%, it may be less than that. About 20% would be preferred, more preferably at least about 30% would be desired. Higher percentages would be especially preferred.

2. <u>Example of four letter alphabet</u>

A four letter alphabet may be conceptualized in at least two different ways from the two letter alphabet. One way, is to consider the four possible values at each position and to analogize in a similar fashion to the binary example each of the overlaps. A second way is to group the binary digits into groups.

Using the first means, the overlap comparisons are performed with a four letter alphabet rather than a two letter alphabet. Then, in contrast to the binary system with 10 positions where 2^{10} = 1024 possible sequences, in a 4-character alphabet with 10 positions, there will actually be 4^{10} =

1,048,576 possible sequences. Thus, the complexity of a four character sequence has a much larger number of possible sequences compared to a two character sequence. Note, however, that there are still only 6 different internal 5-mers. For simplicity, we shall examine a 5 character string with 3 character subsequences. Instead of only 1 and 0, the characters may be designated, e.g., A, C, G, and T. Let us take the sequence GGCTA. The 3-mer subsequences are:

GGC
GCT
CTA

Given these subsequences, there is one sequence, or at most only a few sequences which would produce that combination of subsequences, i.e., GGCTA.

Alternatively, with a four character universe, the binary system can be looked at in pairs of digits. The pairs would be 00, 01, 10, and 11. In this manner, the earlier used sequence 1010011100 is looked at as 10,10,01,11,00. Then the first character of two digits is selected from the possible universe of the four representations 00, 01, 10, and 11. Then a probe would be in an even number of digits, e.g., not five digits, but, three pairs of digits or six digits. A similar comparison is performed and the possible overlaps determined. The 3-pair subsequences are:

10,10,01
10,01,11
01,11,00

and the overlap reconstruction produces 10,10,01,11,00.

The latter of the two conceptual views of the 4 letter alphabet provides a representation which is similar to what would be provided in a digital computer. The applicability to a four nucleotide alphabet is easily seen by assigning, e.g., 00 to A, 01 to C, 10 to G, and 11 to T. And, in fact, if such a correspondence is used, both examples for the 4 character sequences can be seen to represent the same target sequence. The applicability of the hybridization method and its analysis for determining the ultimate sequence is easily seen if A is the representation of adenine, C is the

representation of cytosine, G is the representation of guanine, and T is the representation of thymine or uracil.

3. Generalization to m-letter Alphabet

This reconstruction process may be applied to polymers of virtually any number of possible characters in the alphabet, and for virtually any length sequence to be sequenced, though limitations, as discussed below, will limit its efficiency at various extremes of length. It will be recognized that the theory can be applied to a large diversity of systems where sequence is important.

For example, the method could be applied to sequencing of a polypeptide. A polypeptide can have any of twenty natural amino acid possibilities at each position. A twenty letter alphabet is amenable to sequencing by this method so long as reagents exist for recognizing shorter subsequences therein. A preferred reagent for achieving that goal would be a set of monoclonal antibodies each of which recognizes a specific three contiguous amino acid subsequence. A complete set of antibodies which recognize all possible subsequences of a given length, e.g., 3 amino acids, and preferably with a uniform affinity, would be $20^3 = 8000$ reagents.

It will also be recognized that each target sequence which is recognized by the specific reagents need not have homogeneous termini. Thus, fragments of the entire target sequence will also be useful for hybridizing appropriate subsequences. It is, however, preferable that there not be a significant amount of labeled homogeneous contaminating extraneous sequences. This constraint does usually require the purification of the target molecule to be sequenced, but a specific label technique would dispense with a purification requirement if the unlabeled extraneous sequences do not interfere with the labeled sequences.

In addition, conformational effects of target polypeptide folding may, in certain embodiments, be negligible if the polypeptide is fragmented into sufficiently small peptides, or if the interaction is performed under conditions where conformation, but not specific interaction, is disrupted.

B. Complications

Two obvious complications exist with the method of sequence analysis by hybridization. The first results from a probe of inappropriate length while the second relates to internally repeated sequences.

The first obvious complication is a problem which arises from an inappropriate length of recognition sequence, which causes problems with the specificity of recognition. For example, if the recognized sequence is too short, every sequence which is utilized will be recognized by every probe sequence. This occurs, e.g., in a binary system where the probes are each of sequences which occur relatively frequently, e.g., a two character probe for the binary system. Each possible two character probe would be expected to appear ¼ of the time in every single two character position. Thus, the above sequence example would be recognized by each of the 00, 10, 01, and 11. Thus, the sequence information is virtually lost because the resolution is too low and each recognition reagent specifically binds at multiple sites on the target sequence.

The number of different probes which bind to a target depends on the relationship between the probe length and the target length. At the extreme of short probe length, the just mentioned problem exists of excessive redundancy and lack of resolution. The lack of stability in recognition will also be a problem with extremely short probes. At the extreme of long probe length, each entire probe sequence is on a different position of a substrate. However, a problem arises from the number of possible sequences, which goes up dramatically with the length of the sequence. Also, the specificity of recognition begins to decrease as the contribution to binding by any particular subunit may become sufficiently low that the system fails to distinguish the fidelity of recognition. Mismatched hybridization may be a problem with the polynucleotide sequencing applications, though the fingerprinting and mapping applications may not be so strict in their fidelity requirements. As indicated above, a thirty

position binary sequence has over a million possible sequences, a number which starts to become unreasonably large in its required number of different sequences, even though the target length is still very short. Preparing a substrate with all sequence possibilities for a long target may be extremely difficult due to the many different oligomers which must be synthesized.

The above example illustrates how a long target sequence may be reconstructed with a reasonably small number of shorter subsequences. Since the present day resolution of the regions of the substrate having defined oligomer probes attached to the substrate approaches about 10 microns by 10 microns for resolvable regions, about 10^6, or 1 million, positions can be placed on a one centimeter square substrate. However, high resolution systems may have particular disadvantages which may be outweighed using the lower density substrate matrix pattern. For this reason, a sufficiently large number of probe sequences can be utilized so that any given target sequence may be determined by hybridization to a relatively small number of probes.

A second complication relates to convergence of sequences to a single subsequence. This will occur when a particular subsequence is repeated in the target sequence. This problem can be addressed in at least two different ways. The first, and simpler way, is to separate the repeat sequences onto two different targets. Thus, each single target will not have the repeated sequence and can be analyzed to its end. This solution, however, complicates the analysis by requiring that some means for cutting at a site between the repeats can be located. Typically a careful sequencer would want to have two intermediate cut points so that the intermediate region can also be sequenced in both directions across each of the cut points. This problem is inherent in the hybridization method for sequencing but can be minimized by using a longer known probe sequence so that the frequency of probe repeats is decreased.

Knowing the sequence of flanking sequences of the repeat will simplify the use of polymerase chain reaction (PCR)

or a similar technique to further definitively determine the sequence between sequence repeats. Probes can be made to hybridize to those known sequences adjacent the repeat sequences, thereby producing new target sequences for analysis. See, e.g., Innis et al. (eds.) (1990) PCR Protocols: A Guide to Methods and Applications, Academic Press; and methods for synthesis of oligonucleotide probes, see, e.g., Gait (1984) Oligonucleotide Synthesis: A Practical Approach, IRL Press, Oxford.

Other means for dealing with convergence problems include using particular longer probes, and using degeneracy reducing analogues, see, e.g., Macevicz, S. (1990) PCT publication number WO 90/04652, which is hereby incorporated herein by reference. By use of stretches of the degeneracy reducing analogues with other probes in particular combinations, the number of probes necessary to fully saturate the possible oligomer probes is decreased. For example, with a stretch of 12-mers having the central 4-mer of degenerate nucleotides, in combination with all of the possible 8-mers, the collection numbers twice the number of possible 8-mers, e.g. 65,536 + 65,536 = 131,072, but the population provides screening equivalent to all possible 12-mers.

By way of further explanation, all possible oligonucleotide 8-mers may be depicted in the fashion:

N1-N2-N3-N4-N5-N6-N7-N8,

in which there are 4^8 = 65,536 possible 8-mers. As described in U.S.S.N. __/___,___, attorney docket number 11509-28 (automated VLSIPS), producing all possible 8-mers requires 4x8 = 32 chemical binary synthesis steps to produce the entire matrix pattern of 65,536 8-mer possibilities. By incorporating degeneracy reducing nucleotides, D's, which hybridize nonselectively to any corresponding complementary nucleotide, new oligonucleotides 12-mers can be made in the fashion:

N1-N2-N3-N4-D-D-D-D-N5-N6-N7-N8,

in which there are again, as above, only 4^8 = 65,536 possible "12-mers", which in reality only have 8 different nucleotides.

However, it can be seen that each possible 12-mer probe could be represented by a group of the two 8-mer types.

Moreover, repeats of less than 12 nucleotides would not converge, or cause repeat problems in the analysis. Thus, instead of requiring a collection of probes corresponding to all 12-mers, or 4^{12} = 16,777,216 different 12-mers, the same information can be derived by making 2 sets of "8-mers" consisting of the typical 8-mer collection of 4^8 = 65,536 and the "12-mer" set with the degeneracy reducing analogues, also requiring making 4^8 = 65,536. The combination of the two sets, requires making 65,536 + 65,536 = 131,072 different molecules, but giving the information of 16,777,216 molecules. Thus, incorporating the degeneracy reducing analogue decreases the number of molecules necessary to get 12-mer resolution by a factor of about 128-fold.

C. Non-polynucleotide Embodiments

The above example is directed towards a polynucleotide embodiment. This application is relatively easily achieved because the specific reagents will typically be complementary oligonucleotides, although in certain embodiments other specific reagents may be desired. For example, there may be circumstances where other than complementary base pairing will be utilized. The polynucleotide targets, will usually be single strand, but may be double or triple stranded in various applications. However, a triple stranded specific interaction might be sometimes desired, or a protein or other specific binding molecule may be utilized. For example, various promoter or DNA sequence specific binding proteins might be used, including, e.g., restriction enzyme binding domains, other binding domains, and antibodies. Thus, specific recognition reagents besides oligonucleotides may be utilized.

For other polymer targets, the specific reagents will often be polypeptides. These polypeptides may be protein binding domains from enzymes or other proteins which display specificity for binding. Usually an antibody molecule may be used, and monoclonal antibodies may be particularly desired. Classical methods may be applied for preparing antibodies, see, e.g., Harlow and Lane (1988) Antibodies: A Laboratory Manual Cold Spring Harbor Press, New York; and Goding (1986)

Monoclonal Antibodies: Principles and Practice (2d Ed.) Academic Press, San Diego. Other suitable techniques for in vitro exposure of lymphocytes to the antigens or selection of libraries of antibody binding sites are described, e.g., in Huse et al. (1989) Science 246:1275-1281; and Ward et al. 91989) Nature 341:544-546, each of which is hereby incorporated herein by reference. Unusual antibody production methods are also described, e.g., in Hendricks et al. (1989) BioTechnology 7:1271-1274; and Hiatt et al. (1989) Nature 342:76-78, each of which is hereby incorporated herein by reference. Other molecules which may exhibit specific binding interaction may be useful for attachment to a VLSIPS substrate by various methods, including the caged biotin methods, see, e.g., U.S.S.N. 07/435,316 (caged biotin parent), and U.S.S.N. 07/612,671 (caged biotin CIP).

The antibody specific reagents should be particularly useful for the polypeptide, carbohydrate, and synthetic polymer applications. Individual specific reagents might be generated by an automated process to generate the number of reagents necessary to advantageously use the high density positional matrix pattern. In an alternative approach, a plurality of hybridoma cells may be screened for their ability to bind to a VLSIPS matrix possessing the desired sequences whose binding specificity is desired. Each cell might be individually grown up and its binding specificity determined by VLSIPS apparatus and technology. An alternative strategy would be to expose the same VLSIPS matrix to a polyclonal serum of high titer. By a successively large volume of serum and different animals, each region of the VLSIPS substrate would have attached to it a substantial number of antibody molecules with specificity of binding. The substrate, with non-covalently bound antibodies could be derivatized and the antibodies transferred to an adjacent second substrate in the matrix pattern in which the antibody molecules had attached to the first matrix. If the sensitivity of detection of binding interaction is sufficiently high, such a low efficiency transfer of antibody molecules may produce a sufficiently high signal to be useful for many purposes, including the sequencing applications.

In another embodiment, capillary forces may be used to transfer the selected reagents to a new matrix, to which the reagents would be positionally attached in the pattern of the recognized sequences. Or, the reagents could be transversely electrophoresed, magnetically transferred, or otherwise transported to a new substrate in their retained positional pattern.

III. POLYNUCLEOTIDE SEQUENCING

In principle, the making of a substrate having a positionally defined matrix pattern of all possible oligonucleotides of a given length involves a conceptually simple method of synthesizing each and every different possible oligonucleotide, and affixed to a definable position. Oligonucleotide synthesis is presently mechanized and enabled by current technology, see, e.g., U.S.S.N. 07/362,901 (VLSIPS parent); U.S.S.N. 07/492,462 (VLSIPS CIP); and instruments supplied by Applied Biosystems, Foster City, California.

A. Preparation of Substrate Matrix

The production of the collection of specific oligonucleotides used in polynucleotide sequencing may be produced in at least two different ways. Present technology certainly allows production of ten nucleotide oligomers on a solid phase or other synthesizing system. See, e.g., instrumentation provided by Applied Biosystems, Foster City, California. Although a single oligonucleotide can be relatively easily made, a large collection of them would typically require a fairly large amount of time and investment. For example, there are 4^{10} = 1,048,576 possible ten nucleotide oligomers. Present technology allows making each and every one of them in a separate purified form though such might be costly and laborious.

Once the desired repertoire of possible oligomer sequences of a given length have been synthesized, this collection of reagents may be individually positionally attached to a substrate, thereby allowing a batchwise hybridization step. Present technology also would allow the

possibility of attaching each and every one of these 10-mers to a separate specific position on a solid matrix. This attachment could be automated in any of a number of ways, particularly use of a caged biotin type linking. This would produce a matrix having each of different possible 10-mers.

A batchwise hybridization is much preferred because of its reproducibility and simplicity. An automated process of attaching various reagents to positionally defined sites on a substrate is provided in U.S.S.N. 07/492,462 (VLSIPS CIP); U.S.S.N. __/___,___, attorney docket number 11509-28 (automated VLSIPS); and U.S.S.N. 07/612,671 (caged biotin CIP); each of which is hereby incorporated herein by reference.

Instead of separate synthesis of each oligonucleotide, these oligonucleotides are conveniently synthesized in parallel by sequential synthetic processes on a defined matrix pattern as provided in U.S.S.N. 07/492,462 (VLSIPS CIP); and U.S.S.N. __/___,___, attorney docket number 11509-28 (automated VLSIPS), which are incorporated herein by reference. Here, the oligonucleotides are synthesized stepwise on a substrate at positionally separate and defined positions. Use of photosensitive blocking reagents allows for defined sequences of synthetic steps over the surface of a matrix pattern. By use of the binary masking strategy, the surface of the substrate can be positioned to generate a desired pattern of regions, each having a defined sequence oligonucleotide synthesized and immobilized thereto.

Although the prior art technology can be used to generate the desired repertoire of oligonucleotide probes, an efficient and cost effective means would be to use the VLSIPS technology described in U.S.S.N. 07/492,462 (VLSIPS CIP) and U.S.S.N. __/___,___, attorney docket number 11509-28 (automated VLSIPS). In this embodiment, the photosensitive reagents involved in the production of such a matrix are described below.

The regions for synthesis may be very small, usually less than about 100 μm x 100 μm, more usually less than about 50 μm x 50 μm. The photolithography technology allows synthetic regions of less than about 10 μm x 10 μm, about

3 μm x 3 μm, or less. The detection also may detect such sized regions, though larger areas are more easily and reliably measured.

At a size of about 30 microns by 30 microns, one million regions would take about 11 centimeters square or a single wafer of about 4 centimeters by 4 centimeters. Thus the present technology provides for making a single matrix of that size having all one million plus possible oligonucleotides. Region size are sufficiently small to correspond to densities of at least about 5 regions/cm^2, 20 regions/cm^2, 50 regions/cm^2, 100 regions/cm^2, and greater, including 300 regions/cm^2, 1000 regions/cm^2, 3K regions/cm^2, 10K regions/cm^2, 30K regions/cm^2, 100K regions/cm^2, 300K regions/cm^2 or more, even in excess of one million regions/cm^2.

Although the pattern of the regions which contain specific sequences is theoretically not important, for practical reasons certain patterns will be preferred in synthesizing the oligonucleotides. The application of binary masking algorithms for generating the pattern of known oligonucleotide probes is described in related U.S.S.N. __/___,___, attorney docket number 11509-28 (automated VLSIPS) which was filed simultaneously with this application. By use of these binary masks, a highly efficient means is provided for producing the substrate with the desired matrix pattern of different sequences. Although the binary masking strategy allows for the synthesis of all lengths of polymers, the strategy may be easily modified to provide only polymers of a given length. This is achieved by omitting steps where a subunit is not attached.

The strategy for generating a specific pattern may take any of a number of different approaches. These approaches are well described in related application U.S.S.N. __/___,___, attorney docket number 11509-28 (automated VLSIPS) and include a number of binary masking approaches which will not be exhaustively discussed herein. However, the binary masking and binary synthesis approaches provide a maximum of diversity with a minimum number of actual synthetic steps.

The length of oligonucleotides used in sequencing applications will be selected on criteria determined to some extent by the practical limits discussed above. For example, if probes are made as oligonucleotides, there will be 65,536 possible eight nucleotide sequences. If a nine subunit oligonucleotide is selected, there are 262,144 possible permeations of sequences. If a ten-mer oligonucleotide is selected, there are 1,048,576 possible permeations of sequences. As the number gets larger, the required number of positionally defined subunits necessary to saturate the possibilities also increases. With respect to hybridization conditions, the length of the matching necessary to converse stability of the conditions selected can be compensated for. See, e.g., Kanehisa, M. (1984) Nuc. Acids Res. 12:203-213, which is hereby incorporated herein by reference.

Although not described in detail here, but below for oligonucleotide probes, the VLSIPS technology would typically use a photosensitive protective group on an oligonucleotide. Sample oligonucleotides are shown in Figure 1. In particular, the photoprotective group on the nucleotide molecules may be selected from a wide variety of positive light reactive groups preferably including nitro aromatic compounds such as o-nitro-benzyl derivatives or benzylsulfonyl. See, e.g., Gait (1984) Oligonucleotide Synthesis: A Practical Approach, IRL Press, Oxford, which is hereby incorporated herein by reference. In a preferred embodiment, 6-nitro-veratryl oxycarbony (NVOC), 2-nitrobenzyl oxycarbonyl (NBOC), or α,α-dimethyl-dimethoxybenzyl oxycarbonyl (DEZ) is used. Photoremovable protective groups are described in, e.g., Patchornik (1970) J. Amer. Chem. Soc. 92:6333-____; and Amit et al. (1974) J. Organic Chem. 39:192-___; each of which is hereby incorporated herein by reference.

A preferred linker for attaching the oligonucleotide to a silicon matrix is illustrated in Figure 2. A more detailed description is provided below. A photosensitive blocked nucleotide may be attached to specific locations of unblocked prior cycles of attachments on the substrate and can be successively built up to the correct length oligonucleotide probe.

It should be noted that multiple substrates may be simultaneously exposed to a single target sequence where each substrate is a duplicate of one another or where, in combination, multiple substrates together provide the complete or desired subset of possible subsequences. This provides the opportunity to overcome a limitation of the density of positions on a single substrate by using multiple substrates. In the extreme case, each probe might be attached to a single bead or substrate and the beads sorted by whether there is a binding interaction. Those beads which do bind might be encoded to indicate the subsequence specificity of reagents attached thereto.

Then, the target may be bound to the whole collection of beads and those beads that have appropriate specific reagents on them will bind to target. Then a sorting system may be utilized to sort those beads that actually bind the target from those that do not. This may be accomplished by presently available cell sorting devices or a similar apparatus. After the relatively small number of beads which have bound the target have been collected, the encoding scheme may be read off to determine the specificity of the reagent on the bead. An encoding system may include a magnetic system, a shape encoding system, a color encoding system, or a combination of any of these, or any other encoding system. Once again, with the collection of specific interactions that have occurred, the binding may be analyzed for sequence information, fingerprint information, or mapping information.

The parameters of polynucleotide sizes of both the probes and target sequences are determined by the applications and other circumstances. The length of the oligonucleotide probes used will depend in part upon the limitations of the VLSIPS technology to provide the number of desired probes. For example, in an absolute sequencing application, it is often useful to have virtually all of the possible oligonucleotides of a given length. As indicated above, there are 65,536 8-mers, 262,144 9-mers, 1,048,576 10-mers, 4,194,304 11-mers, etc. As the length of the oligomer increases the number of different probes which must be synthesized also increases at a

rate of a factor of 4 for every additional nucleotide. Eventually the size of the matrix and the limitations in the resolution of regions in the matrix will reach the point where an increase in number of probes becomes disadvantageous. However, this sequencing procedure requires that the system be able to distinguish, by appropriate selection of hybridization and washing conditions, between binding of absolute fidelity and binding of complementary sequences containing mismatches. On the other hand, if the fidelity is unnecessary, this discrimination is also unnecessary and a significantly longer probe may be used. Significantly longer probes would typically be useful in fingerprinting or mapping applications.

The length of the probe is selected for a length that it will bind with specificity to possible targets. The hybridization conditions are also very important in that they will determine how close the homology of complementary binding will be detected. In fact, a single target may be evaluated at a number of different conditions to determine its spectrum of specificity for binding particular probes. This may find use in a number of other applications besides the polynucleotide sequencing fingerprinting or mapping. For example, it will be desired to determine the spectrum of binding affinities and specificities of cell surface antigens with binding by particular antibodies immobilized on the substrate surface, particularly under different interaction conditions. In a related fashion, different regions with reagents having differing affinities or levels of specificity may allow such a spectrum to be defined using a single incubation, where various regions, at a given hybridization condition, show the binding affinity. For example, fingerprint probes of various lengths, or with specific defined non-matches may be used. Unnatural nucleotides or nucleotides exhibiting modified specificity of complementary binding are described in greater detail in Macevicz (1990) PCT pub. No. WO 90/04652; and see the section on modified nucleotides in the Sigma Chemical Company catalogue.

B. Labeling Target Nucleotide

The label used to detect the target sequences will be determined, in part, by the detection methods being applied. Thus, the labeling method and label used are selected in combination with the actual detecting systems being used.

Once a particular label has been selected, appropriate labeling protocols will be applied, as described below for specific embodiments. Standard labeling protocols for nucleic acids are described, e.g., in Sambrook et al.; Kambara, H. et al. (1988) BioTechnology 6:816-821; Smith, L. et al. (1985) Nuc. Acids Res. 13:2399-2412; for polypeptides, see, e.g., Allen G. (1989) Sequencing of Proteins and Peptides, Elsevier, New York, especially chapter 5, and Greenstein and Winitz (1961) Chemistry of the Amino Acids, Wiley and Sons, New York. Carbohydrate labeling is described, e.g., in Chaplin and Kennedy (1986) Carbohydrate Analysis: A Practical Approach, IRL Press, Oxford. Labeling of other polymers will be performed by methods applicable to them as recognized by a person having ordinary skill in manipulating the corresponding polymer.

In some embodiments, the target need not actually be labeled if a means for detecting where interaction takes place is available. As described below, for a nucleic acid embodiment, such may be provided by an intercalating dye which intercalates only into double stranded segments, e.g., where interaction occurs. See, e.g., Sheldon et al. U.S. Pat. No. 4,582,789.

In many uses, the target sequence will be absolutely homogeneous, both with respect to the total sequence and with respect to the ends of each molecule. Homogeneity with respect to sequence is important to avoid ambiguity. It is preferable that the target sequences of interest not be contaminated with a significant amount of labeled contaminating sequences. The extent of allowable contamination will depend on the sensitivity of the detection system and the inherent signal to noise of the system. Homogeneous contamination sequences will be particularly disruptive of the sequencing procedure.

However, although the target polynucleotide must have a unique sequence, the target molecules need not have identical ends. In fact, the homogeneous target molecule preparation may be randomly sheared to increase the numerical number of molecules. Since the total information content remains the same, the shearing results only in a higher number of distinct sequences which may be labeled and bind to the probe. This fragmentation may give a vastly superior signal relative to a preparation of the target molecules having homogeneous ends. The signal for the hybridization is likely to be dependent on the numerical frequency of the target-probe interactions. If a sequence is individually found on a larger number of separate molecules a better signal will result. In fact, shearing a homogeneous preparation of the target may often be preferred before the labeling procedure is performed, thereby producing a large number of labeling groups associated with each subsequence.

C. Hybridization Conditions

The hybridization conditions between probe and target should be selected such that the specific recognition interaction, i.e., hybridization, of the two molecules is both sufficiently specific and sufficiently stable. See, e.g., Hames and Higgins (1985) Nucleic Acid Hybridisation: A Practical Approach, IRL Press, Oxford. These conditions will be dependent both on the specific sequence and often on the guanine and cytosine (GC) content of the complementary hybrid strands. The conditions may often be selected to be universally equally stable independent of the specific sequences involved. This typically will make use of a reagent such as an arylammonium buffer. See, Wood et al. (1985) "Base Composition-independent Hybridization in Tetramethylammonium Chloride: A Method for Oligonucleotide Screening of Highly Complex Gene Libraries," Proc. Natl. Acad. Sci. USA, 82:1585-1588; and Krupov et al. (1989) "An Oligonucleotide Hybridization Approach to DNA Sequencing," FEBS Letters, 256:118-122; each of which is hereby incorporated herein by reference. An arylammonium buffer tends to minimize

differences in hybridization rate and stability due to GC content. By virtue of the fact that sequences then hybridize with approximately equal affinity and stability, there is relatively little bias in strength or kinetics of binding for particular sequences. Temperature and salt conditions along with other buffer parameters should be selected such that the kinetics of renaturation should be essentially independent of the specific target subsequence or oligonucleotide probe involved. In order to ensure this, the hybridization reactions will usually be performed in a single incubation of all the substrate matrices together exposed to the identical same target probe solution under the same conditions.

Alternatively, various substrates may be individually treated differently. Different substrates may be produced, each having reagents which bind to target subsequences with substantially identical stabilities and kinetics of hybridization. For example, all of the high GC content probes could be synthesized on a single substrate which is treated accordingly. In this embodiment, the arylammonium buffers could be unnecessary. Each substrate is then treated in a manner that the collection of substrates show essentially uniform binding and the hybridization data of target binding to the individual substrate matrix is combined with the data from other substrates to derive the necessary subsequence binding information. The hybridization conditions will usually be selected to be sufficiently specific that the fidelity of base matching will be properly discriminated. Of course, control hybridizations should be included to determine the stringency and kinetics of hybridization.

D. <u>Detection; VLSIPS Scanning</u>

The next step of the sequencing process by hybridization involves labeling of target polynucleotide molecules. A quickly and easily detectable signal is preferred. The VLSIPS apparatus is designed to easily detect a fluorescent label, so fluorescent tagging of the target sequence is preferred. Other suitable labels include heavy metal labels, magnetic probes, chromogenic labels (e.g.,

phosphorescent labels, dyes, and fluorophores) spectroscopic labels, enzyme linked labels, radioactive labels, and labeled binding proteins. Additional labels are described in U.S. Pat. No. 4,366,241, which is incorporated herein by reference.

The detection methods used to determine where hybridization has taken place will typically depend upon the label selected above. Thus, for a fluorescent label a fluorescent detection step will typically be used. U.S.S.N. 07/492,462 (VLSIPS CIP) and U.S.S.N. __/___,___, attorney docket number 11509-28 (automated VLSIPS) describe apparatus and mechanisms for scanning a substrate matrix using fluorescence detection, but a similar apparatus is adaptable for other optically detectable labels.

The detection method provides a positional localization of the region where hybridization has taken place. However, the position is correlated with the specific sequence of the probe since the probe has specifically been attached or synthesized at a defined substrate matrix position. Having collected all of the data indicating the subsequences present in the target sequence, this data may be aligned by overlap to reconstruct the entire sequence of the target, as illustrated above.

It is also possible to dispense with actual labeling if some means for detecting the positions of interaction between the sequence specific reagent and the target molecule are available. This may take the form of an additional reagent which can indicate the sites either of interaction, or the sites of lack of interaction, e.g., a negative label. For the nucleic acid embodiments, locations of double strand interaction may be detected by the incorporation of intercalating dyes, or other reagents such as antibody or other reagents that recognize helix formation, see, e.g., Sheldon, et al. (1986) U.S. Pat. No. 4,582,789, which is hereby incorporated herein by reference.

E. Analysis

Although the reconstruction can be performed manually as illustrated above, a computer program will typically be used

to perform the overlap analysis. A program may be written and run on any of a large number of different computer hardware systems. The variety of operating systems and languages useable will be recognized by a computer software engineer. Various different languages may be used, e.g., BASIC; C; PASCAL; etc. A simple flow chart of data analysis is illustrated in Figure 4.

F. Substrate Reuse

Finally, after a particular sequence has been hybridized and the pattern of hybridization analyzed, the matrix substrate should be reusable and readily prepared for exposure to a second or subsequent target polynucleotides. In order to do so, the hybrid duplexes are disrupted and the matrix treated in a way which removes all traces of the original target. The matrix may be treated with various detergents or solvents to which the substrate, the oligonucleotide probes, and the linkages to the substrate are inert. This treatment may include an elevated temperature treatment, treatment with organic or inorganic solvents, modifications in pH, and other means for disrupting specific interaction. Thereafter, a second target may actually be applied to the recycled matrix and analyzed as before.

G. Non-Polynucleotide Aspects

Although the sequencing, fingerprinting, and mapping functions will make use of the natural sequence recognition property of complementary nucleotide sequences, the non-polynucleotide sequences typically require other sequence recognition reagents. These reagents will take the form, typically, of proteins exhibiting binding specificity, e.g., enzyme binding sites or antibody binding sites.

Enzyme binding sites may be derived from promoter proteins, restriction enzymes, and the like. See, e.g., Stryer, L. (1988) Biochemistry, W.H.Freeman, Palo Alto. Antibodies will typically be produced using standard procedures, see, e.g., Harlow and Lane (1988) Antibodies: A Laboratory Manual, Cold Spring Harbor Press, New York; and

Goding (1986) Monoclonal Antibodies: Principles and Practice, (2d Ed.) Academic Press, San Diego.

Typically, an antigen, or collection of antigens are presented to an immune system. This may take the form of synthesized short polymers produced by the VLSIPS technology, or by the other synthetic means, or from isolation of natural products. For example, antigen for the polypeptides may be made by the VLSIPS technology, by standard peptide synthesis, by isolation of natural proteins with or without degradation to shorter segments, or by expression of a collection of short nucleic acids of random or defined sequences. See, e.g., Tuerk and Gold (1990) Science 249:505-510, for generation of a collection of randomly mutagenized oligonucleotides useful for expression.

The antigen or collection is presented to an appropriate immune system, e.g., to a whole animal as in a standard immunization protocol, or to a collection of immune cells or equivalent. In particular, see Ward et al. (1989) Nature 341:544-546; and Huse et al. (1989) Science 246:1275-1281, each of which is hereby incorporated herein by reference.

A large diversity of antibodies will be generated, some of which have specificities for the desired sequences. Antibodies may be purified having the desired sequence specificities by isolating the cells producing them. For example, a VLSIPS substrate with the desired antigens synthesized thereon may be used to isolate cells with cell surface reagents which recognize the antigens. The VLSIPS substrate may be used as an affinity reagent to select and recover the appropriate cells. Antibodies from those cells may be attached to a substrate using the caged biotin methodology, or by attaching a targeting molecule, e.g., an oligonucleotide. Alternatively, the supernatants from antibody producing cells can be easily assayed using a VLSIPS substrate to identify the cells producing the appropriate antibodies.

Although cells may be isolated, specific antibody molecules which perform the sequence recognition will also be sufficient. Preferably populations of antibody with a known specificity can be isolated. Supernatants from a large

population of producing cells may be passed over a VLSIPS substrate to bind to the desired antigens attached to the substrate. When a sufficient density of antibody molecules are attached, they may be removed by an automated process, preferably as antibody populations exhibiting specificity of binding.

In one particular embodiment, a VLSIPS substrate, e.g., with a large plurality of fingerprint antigens attached thereto, is used to isolate antibodies from a supernatant of a population of cells producing antibodies to the antigens. Using the substrate as an affinity reagent, the antibodies will attach to the appropriate positionally defined antigens. The antibodies may be carefully removed therefrom, preferably by an automated system which retains their homogeneous specificities. The isolated antibodies can be attached to a new substrate in a positionally defined matrix pattern.

In a further embodiment, these spatially separated antibodies may be isolated using a specific targeting method for isolation. In this embodiment, a linker molecule which attaches to a particular portion of the antibody, preferably away from the binding site, can be attached to the antibodies. Various reagents will be used, including staphylococcus protein A or antibodies which bind to domains remote from the binding site. Alternatively, the antibodies in the population, before affinity purification, may be derivatized with an appropriate reagent compatible with new VLSIPS synthesis. A preferred reagent is a nucleotide which can serve as a linker to synthetic VLSIPS steps for synthesizing a specific sequence thereon. Then, by successive VLSIPS cycles, each of the antibodies attached to the defined antigen regions can have a defined oligonucleotide synthesized thereon and corresponding in area to the region of the substrate having each antigen attached. These defined oligonucleotides will be useful as targeting reagents to attach those antibodies possessing the same target sequence specificity at defined positions on a new substrate, by virtue of having bound to the antigen region, to a new VLSIPS substrate having the complementary target oligonucleotides positionally located on it. In this fashion,

a VLSIPS substrate having the desired antigens attached thereto can be used to generate a second VLSIPS substrate with positionally defined reagents which recognize those antigens.

The selected antigens will typically be selected to be those which define particular functionalities or properties, so as to be useful for fingerprinting and other uses. They will also be useful for mapping and sequencing embodiments.

IV. FINGERPRINTING

A. General

Many of the procedures and techniques used in the polynucleotide sequencing section are also appropriate for fingerprinting applications. See, e.g., Poustka, et al. (1986) Cold Spring Harbor Symposia on Quant. Biol., vol. LI, 131-139, Cold Spring Harbor Press, New York; which is hereby incorporated herein by reference. The fingerprinting method provided herein is based, in part, upon the ability to positionally localize a large number of different specific probes onto a single substrate. This high density matrix pattern provides the ability to screen for, or detect, a very large number of different sequences simultaneously. In fact, depending upon the hybridization conditions, fingerprinting to the resolution of virtually absolute matching of sequence is possible thereby approaching an absolute sequencing embodiment. And the sequencing embodiment is very useful in identifying the probes useful in further fingerprinting uses. For example, characteristic features of genetic sequences will be identified as being diagnostic of the entire sequence. However, in most embodiments, longer probe and target will be used, and for which slight mismatching may not need to be resolved.

B. Preparation of Substrate Matrix

A collection of specific probes may be produced by either of the methods described above in the section on sequencing. Specific oligonucleotide probes of desired lengths may be individually synthesized on a standard oligonucleotide synthesizer. The length of these probes is limited only by the length of the ability of the synthesizer to continue to

accurately synthesize a molecule. Oligonucleotides or sequence fragments may also be isolated from natural sources. Biological amplification methods may be coupled with synthetic synthesizing procedures such as, e.g., polymerase chain reaction.

In one embodiment, the individually isolated probes may be attached to the matrix at defined positions. These probe reagents may be attached by an automated process making use of the caged biotin methodology described in U.S.S.N. 07/612,671 (caged biotin CIP), or using photochemical reagents, see, e.g., Dattagupta et al. (1985) U.S. Pat. No. 4,542,102 and (1987) U.S. Pat. No. 4,713,326. Each individual purified reagent can be attached individually at specific locations on a substrate.

In another embodiment, the VLSIPS synthesizing technique may be used to synthesize the desired probes at specific positions on a substrate. The probes may be synthesized by successively adding appropriate monomer subunits, e.g., nucleotides, to generate the desired sequences.

In another embodiment, a relatively short specific oligonucleotide is used which serves as a targeting reagent for positionally directing the sequence recognition reagent. For example, the sequence specific reagents having a separate additional sequence recognition segment (usually of a different polymer from the target sequence) can be directed to target oligonucleotides attached to the substrate. By use of non-natural targeting reagents, e.g., unusual nucleotide analogues which pair with other unnatural nucleotide analogues and which do not interfere with natural nucleotide interactions, the natural and non-natural portions can coexist on the same molecule without interfering with their individual functionalities. This can combine both a synthetic and biological production system analogous to the technique for targeting monoclonal antibodies to locations on a VLSIPS substrate at defined positions. Unnatural optical isomers of nucleotides may be useful unnatural reagents subject to similar chemistry, but incapable of interfering with the natural biological polymers. See also, U.S.S.N. __/___,___, attorney

docket number 11509-26 (sequencing by synthesis); which is hereby incorporated herein by reference.

After the separate substrate attached reagents are attached to the targeting segment, the two are crosslinked, thereby permanently attaching them to the substrate. Suitable crosslinking reagents are known, see, e.g., Dattagupta et al. (1985) U.S. Pat. No. 4,542,102 and (1987) "Coupling of nucleic acids to solid support by photochemical methods," U.S. Pat. No. 4,713,326, each of which is hereby incorporated herein by reference. Similar linkages for attachment of proteins to a solid substrate are provided, e.g., in Merrifield (1986) Science 232:341-____, which is hereby incorporated herein by reference.

C. Labeling Target Nucleotides

The labeling procedures used in the sequencing embodiments will also be applicable in the fingerprinting embodiments. However, since the fingerprinting embodiments often will involve relatively large target molecules and relatively short oligonucleotide probes, the amount of signal necessary to incorporate into the target sequence may be less critical than in the sequencing applications. For example, a relatively long target with a relatively small number of labels per molecule may be easily amplified or detected because of the relatively large target molecule size.

In various embodiments, it may be desired to cleave the target into smaller segments as in the sequencing embodiments. The labeling procedures and cleavage techniques described in the sequencing embodiments would usually also be applicable here.

D. Hybridization Conditions

The hybridization conditions used in fingerprinting embodiments will typically be less critical than for the sequencing embodiments. The reason is that the amount of mismatching which may be useful in providing the fingerprinting information would typically be far greater than that necessary in sequencing uses. For example, Southern hybridizations do

not typically distinguish between slightly mismatched sequences. Under these circumstances, important and valuable information may be arrived at with less stringent hybridization conditions while providing valuable fingerprinting information. However, since the entire substrate is typically exposed to the target molecule at one time, the binding affinity of the probes should usually be of approximately comparable levels. For this reason, if oligonucleotide probes are being used, their lengths should be approximately comparable and will be selected to hybridize under conditions which are common for most of the probes on the substrate. Much as in a Southern hybridization, the target and oligonucleotide probes are of lengths typically greater than about 25 nucleotides. Under appropriate hybridization conditions, e.g., typically higher salt and lower temperature, the probes will hybridize irrespective of imperfect complementarity. In fact, with probes of greater than, e.g., about fifty nucleotides, the difference in stability of different sized probes will be relatively minor.

Typically the fingerprinting is merely for probing similarity or homology. Thus, the stringency of hybridization can usually be decreased to fairly low levels. See, e.g., Wetmur and Davidson (1968) "Kinetics of Renaturation of DNA," J. Mol. Biol., 31:349-370; and Kanehisa, M. (1984) Nuc. Acids Res., 12:203-213.

E. Detection; VLSIPS Scanning

Detection methods will be selected which are appropriate for the selected label. The scanning device need not necessarily be digitized or placed into a specific digital database, though such would most likely be done. For example, the analysis in fingerprinting could be photographic. Where a standardized fingerprint substrate matrix is used, the pattern of hybridizations may be spatially unique and may be compared photographically. In this manner, each sample may have a characteristic pattern of interactions and the likelihood of identical patterns will preferably be such low frequency that the fingerprint pattern indeed becomes a characteristic pattern virtually as unique as an individual's fingertip fingerprint.

With a standardized substrate, every individual could be, in theory, uniquely identifiable on the basis of the pattern of hybridizing to the substrate.

Of course, the VLSIPS scanning apparatus may also be useful to generate a digitized version of the fingerprint pattern. In this way, the identification pattern can be provided in a linear string of digits. This sequence could also be used for a standardized identification system providing significant useful medical transferability of specific data. In one embodiment, the probes used are selected to be of sufficiently high resolution to measure the antigens of the major histo compatibility complex, it might even be possible to provide transplantation matching data in a linear stream of data. The fingerprinting data may provide a condensed version, or summary, of the linear genetic data, or any other information data base.

F. Analysis

The analysis of the fingerprint will often be much simpler than a total sequence determination. However, there may be particular types of analysis which will be substantially simplified by a selected group of probes. For example, probes which exhibit particular populational heterogeneity may be selected. In this way, analysis may be simplified and practical utility enhanced merely by careful selection of the specific probes and a careful matrix layout of those probes.

G. Substrate Reuse

As with the sequencing application, the fingerprinting usages may also take advantage of the reusability of the substrate. In this way, the interactions can be disrupted, the substrate treated, and the renewed substrate is equivalent to an unused substrate.

H. Non-polynucleotide Aspects

Besides polynucleotide applications, the fingerprinting analysis may be applied to other polymers, especially polypeptides, carbohydrates, and other polymers,

both organic and inorganic. Besides using the fingerprinting method for analyzing a particular polymer, the fingerprinting method may be used to characterize various samples. For example, a cell or population of cells may be tested for their expression of specific antigens or their mRNA sequence intent. For example, a T-cell may be classified by virtue of its combination of expressed surface antigens. With specific reagents which interact with these antigens, a cell or a population of cells or a lysed cell may be exposed to a VLSIPS substrate. The biological sample may be classified or characterized by analyzing the pattern of specific interaction. This may be applicable to a cell or tissue type, to the expressed messenger RNA population expressed by a cell to the genetic content of a cell, or to virtually any sample which can be classified and/or identified by its combination of specific molecular properties.

The ability to generate a high density means for screening the presence or absence of specific interactions allows for the possibility of screening for, if not saturating, all of a very large number of possible interactions. This is very powerful in providing the means for testing the combinations of molecular properties which can define a class of samples. For example, a species of organism may be characterized by its DNA sequences, e.g., a genetic fingerprint. By using a fingerprinting method, it may be determined that all members of that species are sufficiently similar in specific sequences that they can be easily identified as being within a particular group. Thus, newly defined classes may be resolved by their similarity in fingerprint patterns. Alternatively, a non-member of that group will fail to share those many identifying characteristics. However, since the technology allows testing of a very large number of specific interactions, it also provides the ability to more finely distinguish between closely related different cells or samples. This will have important applications in diagnosing viral, bacterial, and other pathological on nonpathological infections.

In particular, cell classification may be defined by any of a number of different properties. For example, a cell class may be defined by its DNA sequences contained therein. This allows species identification for parasitic or other infections. For example, the human cell is presumably genetically distinguishable from a monkey cell, but different human cells will share many genetic markers. At higher resolution, each individual human genome will exhibit unique sequences that can define it as a single individual.

Likewise, a developmental stage of a cell type may be definable by its pattern of expression of messenger RNA. For example, in particular stages of cells, high levels of ribosomal RNA are found whereas relatively low levels of other types of messenger RNAs may be found. The high resolution distinguishability provided by this fingerprinting method allows the distinction between cells which have relatively minor differences in its expressed mRNA population. Where a pattern is shown to be characteristic of a stage, a stage may be defined by that particular pattern of messenger RNA expression.

In a similar manner, the antigenic determinants found on a protein may very well define the cell class. For example, immunological T-cells are distinguishable from B-cells because, in part, the cell surface antigens on the cell types are distinguishable. Different T-cell subclasses can be also distinguished from one another by whether they contain particular T-cell antigens. The present invention provides the possibility for high resolution testing of many different interactions simultaneously, and the definition of new cell types will be possible.

The high resolution VLSIPS substrate may also be used as a very powerful diagnostic tool to test the combination of presence, of a plurality of different assays from a biological sample. For example, a cancerous condition may be indicated by a combination of various different properties found in the blood. For example, a cancerous condition may be indicated by a combination of expression of various soluble antigens found in the blood along with a high number of various cellular

antigens found on lymphocytes and/or particular cell degradation products. With a substrate as provided herein, a large number of different features can be simultaneously performed on a biological sample. In fact, the high resolution of the test will allow more complete characterization of parameters which define particular diseases. Thus, the power of diagnostic tests may be limited by the extent of statistical correlation with a particular condition rather than with the number of antigens or interactions which are tested. The present invention provides the means to generate this large universe of possible reagents and the ability to actually accumulate that correlative data.

In another embodiment, a substrate as provided herein may be used for genetic screening. This would allow for simultaneous screening of thousands of genetic markers. As the density of the matrix is increased, many more molecules can be simultaneously tested. Genetic screening then becomes a simpler method as the present invention provides the ability to screen for thousands, tens of thousands, and hundreds of thousands, even millions of different possible genetic features. However, the number of high correlation genetic markers for conditions numbers only in the hundreds. Again, the possibility for screening a large number of sequences provides the opportunity for generating the data which can provide correlation between sequences and specific conditions or susceptibility. The present invention provides the means to generate extremely valuable correlations useful for the genetic detection of the causative mutation leading to medical conditions. In still another embodiment, the present invention would be applicable to distinguishing two individuals having identical genetic compositions. The antibody population within an individual is dependent both on genetic and historical factors. Each individual experiences a unique exposure to various infectious agents, and the combined antibody expression is partly determined thereby. Thus, individuals may also be fingerprinted by their immunological content, either of actively expressed antibodies, or their immunological memory. Similar sorts of immunological and environmental histories may

be useful for fingerprinting, perhaps in combination with other screening properties. In particular, the present invention may be useful for screening allergic reactions or susceptibilities, a simple IgE specificity test may be useful in determining a spectrum of allergies.

With the definition of new classes of cells, a cell sorter will be used to purify them. Moreover, new markers for defining that class of cells will be identified. For example, where the class is defined by its RNA content, cells may be screened by antisense probes which detect the presence or absence of specific sequences therein. Alternatively, cell lysates may provide information useful in correlating intracellular properties with extracellular markers which indicate functional differences. Using standard cell sorter technology with a fluorescence or labeled antisense probe which recognizes the internal presence of the specific sequences of interest, the cell sorter will be able to isolate a relatively homogeneous population of cells possessing the particular marker. Using successive probes the sorting process should be able to select for cells having a combination of a large number of different markers.

In a non-polynucleotide embodiment, cells may be defined by the presence of other markers. The markers may be carbohydrates, proteins, or other molecules. Thus, a substrate having particular specific reagents, e.g., antibodies, attached to it should be able to identify cells having particular patterns of marker expression. Of course, combinations of these made be utilized and a cell class may be defined by a combination of its expressed mRNA, its carbohydrate expression, its antigens, and other properties. This fingerprinting should be useful in determining the physiological state of a cell or population of cells.

Having defined a cell type whose function or properties are defined by the reagents attachable to a VLSIPS substrate, such as cellular antigens, these structural manifestations of function may be used to sort cells to generate a relatively homogeneous population of that class of cells. Standard cell sorter technology may be applied to

purify such a population, see, e.g., Dangl, J. and Herzenberg (1982) "Selection of hybridomas and hybridoma variants using the fluorescence activated cell sorter," J. Immunological Methods 52:1-14; and Becton Dickinson, Fluorescence Activated Cell Sorter Division, San Jose, California, and Coulter Diagnostics, Hialeah, Florida.

With the fingerprinted method as in identification means arises from mosaism problems in an organism. A mosaic organism is one whose genetic content in different cells is significantly different. Various clonal populations should have similar genetic fingerprints, though different clonal populations may have different genetic contents. See, for example, Suzuki et al. An Introduction to Genetic Analysis (4th Ed.), Freeman and Co., New York, which is hereby incorporated herein by reference. However, this problem should be a relatively rare problem and could be more carefully evaluated with greater experience using the fingerprinting methods.

The invention will also find use in detecting changes, both genetic and antigenic, e.g., in a rapidly "evolving" protozoa infection, or similarly changing organism.

V. MAPPING

A. General

The use of the present invention for mapping parallels its use for fingerprinting and sequencing. Where a polymer is a linear molecule, the mapping provides the ability to locate particular segments along the length of the polymer. Branched polymers can be treated as a series of individual linear polymers. The mapping provides the ability to locate, in a relative sense, the order of various subsequences. This may be achieved using at least two different approaches.

The first approach is to take the large sequence and fragment it at specific points. The fragments are then ordered and attached to a solid substrate. For example, the clones resulting from a chromosome walking process may be individually attached to the substrate by methods, e.g., caged biotin techniques, indicated earlier. Segments of unknown map position will be exposed to the substrate and will hybridize to

the segment which contains that particular sequence. This procedure allows the rapid determination of a number of different labeled segments, each mapping requiring only a single hybridization step once the substrate is generated. The substrate may be regenerated by removal of the interaction, and the next mapping segment applied.

In an alternative method, a plurality of subsequences can be attached to a substrate. Various short probes may be applied to determine which segments may contain particular overlaps. The theoretical basis and a description of this mapping procedure is contained in, e.g., Evans et al. 1989 "Physical Mapping of Complex Genomes by Cosmid Multiplex Analysis," Proc. Natl. Acad. Sci. USA 86:5030-5034, and other references cited above in the Section labeled "Overall Description." Using this approach, the details of the mapping embodiment are very similar to those used in the fingerprinting embodiment.

B. Preparation of Substrate Matrix

The substrate may be generated in either of the methods generally applicable in the sequencing and fingerprinting embodiments. The substrate may be made either synthetically, or by attaching otherwise purified probes or sequences to the matrix. The probes or sequences may be derived either from synthetic or biological means. As indicated above, the solid phase substrate synthetic methods may be utilized to generate a matrix with positionally defined sequences. In the mapping embodiment, the importance of saturation of all possible subsequences of a preselected length is far less important than in the sequencing embodiment, but the length of the probes used may be desired to be much longer. The processes for making a substrate which has longer oligonucleotide probes should not be significantly different from those described for the sequencing embodiments, but the optimization parameters may be modified to comply with the mapping needs.

C. Labeling

The labeling methods will be similar to those applicable in sequencing and fingerprinting embodiments. Again, the target sequences may be desired to be fragmented.

D. Hybridization/Specific Interaction

The specificity of interaction between the targets and probe would typically be closer to those used for fingerprinting embodiments, where homology is more important than absolute distinguishability of high fidelity complementary hybridization. Usually, the hybridization conditions will be such that merely homologous segments will interact and provide a positive signal. Much like the fingerprinting embodiment, it may be useful to measure the extent of homology by successive incubations at higher stringency conditions. Or, a plurality of different probes, each having various levels of homology may be used. In either way, the spectrum of homologies can be measured.

Where non-nucleic acid hybridization is involved, the specific interactions may also be compared in a fingerprint-like manner. The specific reagents may have less specificity, e.g., monoclonal antibodies which recognize a broader spectrum of sequences may be utilized relative to a sequencing embodiment. Again, the specificity of interaction may be measured under various conditions of increasing stringency to determine the spectrum of matching across the specific probes selected, or a number of different stringency reagents may be included to indicate the binding affinity.

E. Detection

The detection methods used in the mapping procedure will be virtually identical to those used in the fingerprinting embodiment. The detection methods will be selected in combination with the labeling methods.

F. Analysis

The analysis of the data in a mapping embodiment will typically be somewhat different from that in fingerprinting.

The fingerprinting embodiment will test for the presence or absence of specific or homologous segments. However, in the mapping embodiment, the existence of an interaction is coupled with some indication of the location of the interaction. The interaction is mapped in some manner to the physical polymer sequence. Some means for determining the relative positions of different probes is performed. This may be achieved by synthesis of the substrate in pattern, or may result from analysis of sequences after they have been attached to the substrate.

For example, the probes may be randomly positioned at various locations on the substrate. However, the relative positions of the various reagents in the original polymer may be determined by using short fragments, e.g., individually, as target molecules which determine the proximity of different probes. By an automated system of testing each different short fragment of the original polymer, coupled with proper analysis, it will be possible to determine which probes are adjacent one another on the original target sequence and correlate that with positions on the matrix. In this way, the matrix is useful for determining the relative locations of various new segments in the original target molecule. This sort of analysis is described in Evans, and the related references described above.

G. <u>Substrate Reuse</u>

The substrate should be reusable in the manner described in the fingerprinting section. The substrate is renewed by removal of the specific interactions and is washed and prepared for successive cycles of exposure to new target sequences.

H. <u>Non-polynucleotide Aspects</u>

The mapping procedure may be used on other molecules than polynucleotides. Although hybridization is one type of specific interaction which is clearly useful for use in this mapping embodiment, antibody reagents may also be very useful. In the same way that polypeptide sequencing or other polymers may be sequenced by the reagents and techniques described in

the sequencing section and fingerprinting section, the mapping embodiment may also be used similarly.

In another form of mapping, as described above in the fingerprinting section, the developmental map of a cell or biological system may be measured using fingerprinting type technology. Thus, the mapping may be along a temporal dimension rather than along a polymer dimension. The mapping or fingerprinting embodiments may also be used in determining the genetic rearrangements which may be genetically important, as in lymphocyte and B-cell development. In another example, various rearrangements or chromosomal dislocations may be tested by either the fingerprinting or mapping methods. These techniques are similar in many respects and the fingerprinting and mapping embodiments may overlap in many respects.

VI. ADDITIONAL SCREENING AND APPLICATIONS

A. Specific Interactions

As originally indicated in the parent filing of VLSIPS, the production of a high density plurality of spatially segregated polymers provides the ability to generate a very large universe or repertoire of individually and distinct sequence possibilities. As indicated above, particular oligonucleotides may be synthesized in automated fashion at specific locations on a matrix. In fact, these oligonucleotides may be used to direct other molecules to specific locations by linking specific oligonucleotides to other reagents which are in batch exposed to the matrix and hybridized in a complementary fashion to only those locations where the complementary oligonucleotide has been synthesized on the matrix. This allows for spatially attaching a plurality of different reagents onto the matrix instead of individually attaching each separate reagent at each specific location. Although the caged biotin method allows the automated attachment, the speed of the caged biotin attachment process is relatively slow and requires a separate reaction for each reagent being attached. By use of the oligonucleotide method, the specificity of position can be done in an automated and parallel fashion. As each reagent is produced, instead of

directly attaching each reagent at each desired position, the reagent may be attached to a specific desired complementary oligonucleotide which will ultimately be specifically directed toward locations on the matrix having a complementary oligonucleotide attached thereat.

In addition, the technology allows screening for specificity of interaction with particular reagents. For example, the oligonucleotide sequence specificity of binding of a potential reagent may be tested by presenting to the reagent all of the possible subsequences available for binding. Although secondary or higher order sequence specific features might not be easily screenable using this technology, it does provide a convenient, simple, quick, and thorough screen of interactions between a reagent and its target recognition sequences. See, e.g., Pfeifer et al. (1989) Science 246:810-812.

For example, the interaction of a promoter protein with its target binding sequence may be tested for many different, or all, possible binding sequences. By testing the strength of interactions under various different conditions, the interaction of the promoter protein with each of the different potential binding sites may be analyzed. The spectrum of strength of interactions with each different potential binding site may provide significant insight into the types of features which are important in determining specificity.

An additional example of a sequence specific interaction between reagents is the testing of binding of a double stranded nucleic acid structure with a single stranded oligonucleotide. Often, a triple stranded structure is produced which has significant aspects of sequence specificity. Testing of such interactions with either sequences comprising only natural nucleotides, or perhaps the testing of nucleotide analogs may be very important in screening for particularly useful diagnostic or therapeutic reagents. See, e.g., Häner and Dervan (1990) Biochemistry 29:9761-6765, and references therein.

B. Sequence Comparisons

Once a gene is sequenced, the present invention provides means to compare alleles or related sequences to locate and identify differences from the control sequence. This would be extremely useful in further analysis of genetic variability at a specific gene locus.

C. Categorizations

As indicated above in the fingerprinting and mapping embodiments, the present invention is also useful to define specific stages in the temporal sequence of cells, e.g., development, and the resulting tissues within an organism. For example, the developmental stage of a cell, or population of cells, can be dependent upon the expression of particular messenger RNAs or cellular antigens. The screening procedures provided allow for high resolution definition of new classes of cells. In addition, the temporal development of particular cells will be characterized by the presence or expression of various mRNAs. Means to simultaneously screen a plurality or very large number of different sequences as provided. The combination of different markers made available dramatically increases the ability to distinguish fairly closely related cell types. Other markers may be combined with markers and methods made available herein to define new classifications of biological samples, e.g., based upon new combinations of markers.

The presence or absence of particular marker sequences will be used to define temporal developmental stages. Once the stages are defined, fairly simple methods can be applied to actually purify those particular cells. For example, antisense probes or recognition reagents may be used with a cell sorter to select those cells containing or expressing the critical markers. Alternatively, the expression of those sequences may result in specific antigens which may also be used in defining cell classes and sorting those cells away from others. In this way, for example, it should be possible to select a class of omnipotent immune system cells which are able to completely regenerate a human immune system.

Based upon the cellular classes defined by the parameters made available by this technology, purified classes of cells having identifiable differences, structural or functional, are made available.

In an alternative embodiment, a plurality of antigens or specific binding proteins attached to the substrate may be used to define particular cell types. For example, subclasses of T-cells are defined, in part, upon the combination of expressed cell surface antigens. The present invention allows for the simultaneous screening of a large plurality of different antigens together. Thus, higher resolution classification of different T-cell subclasses becomes possible and, with the definitions and functional differences which correlate with those antigenic or other parameters, the ability to purify those cell types becomes available. This is applicable not only to T-cells, lymphocyte cells, or even to freely circulating cells. Many of the cells for which this would be most useful will be immobile cells found in particular tissues or organs. Tumor cells will be diagnosed or detected using these fingerprinting techniques. Coupled with a temporal change in structure, developmental classes may also be selected and defined using these technologies. The present invention also provides the ability not only to define new classes of cells based upon functional or structural differences, but it also provides the ability to select or purify populations of cells which share these particular properties. Standard cell sorting procedures using antibody markers may be used to detect extracellular features. Intracellular features would also be amendable by introducing the label reagents into the cell. In particular, antisense DNA or RNA molecules may be introduced into a cell to detect RNA sequences therein. See, e.g., Weintraub (1990) Scientific American 262:40-46.

D. Statistical Correlations

In an additional embodiment, the present invention also allows for the high resolution correlation of medical conditions with various different markers. For example, the present technology, when applied to amniocentesis or other

genetic screening methods, typically screen for tens of different markers at most. The present invention allows simultaneous screening for tens, hundreds, thousands, tens of thousands, hundreds of thousands, and even millions of different genetic sequences. Thus, applying the fingerprinting methods of the present invention to a sufficiently large population allows detailed statistical analysis to be made, thereby correlating particular medical conditions with particular markers, typically antigenic or genetic. Tumor specific antigens will be identified using the present invention.

Various medical conditions may be correlated against an enormous data base of the sequences within an individual. Genetic propensities and correlations then become available and high resolution genetic predictability and correlation become much more easily performed. With the enormous data base, the reliability of the predictions also is better tested. Particular markers which are partially diagnostic of particular medical conditions or medical susceptibilities will be identified and provide direction in further studies and more careful analysis of the markers involved. Of course, as indicated above in the sequencing embodiment, the present invention will find much use in intense sequencing projects. For example, sequencing of the entire human genome in the human genome project will be greatly simplified and enabled by the present invention.

VI. FORMATION OF SUBSTRATE

The substrate is provided with a pattern of specific reagents which are positionally localized on the surface of the substrate. This matrix of positions is defined by the automated system which produces the substrate. The instrument will typically be one similar to that described in U.S.S.N. 07/492,462 (VLSIPS CIP), and U.S.S.N. __/___,___, attorney docket number 11509-28 (automated VLSIPS). The instrumentation described therein is directly applicable to the applications used here. In particular, the apparatus comprises a substrate, typically a silicon containing substrate, on which positions on

the surface may be defined by a coordinate system of positions. These positions can be individually addressed or detected by the VLSIPS apparatus.

Typically, the VLSIPS apparatus uses optical methods used in semiconductor fabrication applications. In this way, masks may be used to photo-activate positions for attachment or synthesis of specific sequences on the substrate. These manipulations may be automated by the types of apparatus described in U.S.S.N. 07/462,492 (VLSIPS CIP) and U.S.S.N. __/___,___, attorney docket number 11509-28 (automated VLSIPS).

Selectively removable protecting groups allow creation of well defined areas of substrate surface having differing reactivities. Preferably, the protecting groups are selectively removed from the surface by applying a specific activator, such as electromagnetic radiation of a specific wavelength and intensity. More preferably, the specific activator exposes selected areas of surface to remove the protecting groups in the exposed areas.

Protecting groups of the present invention are used in conjunction with solid phase oligomer syntheses, such as peptide syntheses using natural or unnatural amino acids, nucleotide syntheses using deoxyribonucleic and ribonucleic acids, oligosaccharide syntheses, and the like. In addition to protecting the substrate surface from unwanted reaction, the protecting groups block a reactive end of the monomer to prevent self-polymerization. For instance, attachment of a protecting group to the amino terminus of an activated amino acid, such as the N-hydroxysuccinimide-activated ester of the amino acid prevents the amino terminus of one monomer from reacting with the activated ester portion of another during peptide synthesis.

Alternatively, the protecting group may be attached to the carboxyl group of an amino acid to prevent reaction at this site. Most protecting groups can be attached to either the amino or the carboxyl group of an amino acid, and the nature of the chemical synthesis will dictate which reactive group will require a protecting group. Analogously, attachment of a protecting group to the 5'-hydroxyl group of a nucleoside

during synthesis using for example, phosphate-triester coupling chemistry, prevents the 5'-hydroxyl of one nucleoside from reacting with the 3'-activated phosphate-triester of another.

Regardless of the specific use, protecting groups are employed to protect a moiety on a molecule from reacting with another reagent. Protecting groups of the present invention have the following characteristics: they prevent selected reagents from modifying the group to which they are attached; they are stable (that is, they remain attached) to the synthesis reaction conditions; they are removable under conditions that do not adversely affect the remaining structure; and once removed, do not react appreciably with the surface or surface-bound oligomer. The selection of a suitable protecting group will depend, of course, on the chemical nature of the monomer unit and oligomer, as well as the specific reagents they are to protect against.

In a preferred embodiment, the protecting groups will be photoactivatable. The properties and uses of photoreactive protecting compounds have been reviewed. See, McCray et al., Ann. Rev. of Biophys. and Biophys. Chem. (1989) 18:239-270, which is incorporated herein by reference. Preferably, the photosensitive protecting groups will be removable by radiation in the ultraviolet (UV) or visible portion of the electromagnetic spectrum. More preferably, the protecting groups will be removable by radiation in the near UV or visible portion of the spectrum. In some embodiments, however, activation may be performed by other methods such as localized heating, electron beam lithography, laser pumping, oxidation or reduction with microelectrodes, and the like. Sulfonyl compounds are suitable reactive groups for electron beam lithography. Oxidative or reductive removal is accomplished by exposure of the protecting group to an electric current source, preferably using microelectrodes directed to the predefined regions of the surface which are desired for activation. A more detailed description of these protective groups is provided in U.S.S.N. __/___,___, attorney docket number 11509-28 (automated VLSIPS), which is hereby incorporated herein by reference.

The density of reagents attached to a silicon substrate may be varied by standard procedures. The surface area for attachment of reagents may be increased by modifying the silicon surface. For example, a matte surface may be machined or etched on the substrate to provide more sites for attachment of the particular reagents. Another way to increase the density of reagent binding sites is to increase the derivitization density of the silicon. Standard procedures for achieving this are described, below.

One method to control the derivatization density is to highly derivatize the substrate with photochemical groups at high density. The substrate is then photolyzed for various predetermined times, which photoactivate the groups at a measurable rate, and react then with a capping reagent. By this method, the density of linker groups may be modulated by using a desired time and intensity of photoactivation.

In many applications, the number of different sequences which may be provided may be limited by the density and the size of the substrate on which the matrix pattern is generated. In situations where the density is insufficiently high to allow the screening of the desired number of sequences, multiple substrates may be used to increase the number of sequences tested. Thus, the number of sequences tested may be increased by using a plurality of different substrates. Because the VLSIPS apparatus is almost fully automated, increasing the number of substrates does not lead to a significant increase in the number of manipulations which must be performed by humans. This again leads to greater reproducibility and speed in the handling of these multiple substrates.

A. Instrumentation

The concept of using VLSIPS generally allows a pattern or a matrix of reagents to be generated. The procedure for making the pattern is performed by any of a number of different methods. An apparatus and instrumentation useful for generating a high density VLSIPS substrate is described in

detail in U.S.S.N. 07/492,462 (VLSIPS CIP) and U.S.S.N. __/___,___, attorney docket number 11509-28 (automated VLSIPS).

B. Binary Masking

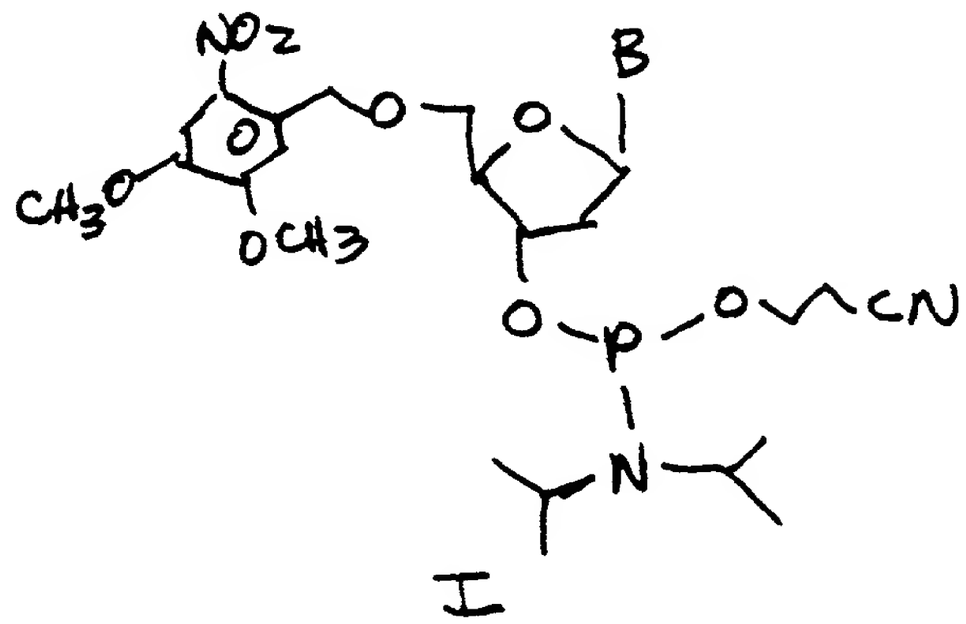
The details of the binary masking are described in an accompanying application filed simultaneously with this, U.S.S.N. __/___,___, attorney docket number 11509-28 (automated VLSIPS) whose specification is incorporated herein by reference.

For example, the binary masking technique allows for producing a plurality of sequences based on the selection of either of two possibilities at any particular location. By a series of binary masking steps, the binary decision may be the determination, on a particular synthetic cycle, whether or not to add any particular one of the possible subunits. By treating various regions of the matrix pattern in parallel, the binary masking strategy provides the ability to carry out spatially addressable parallel synthesis.

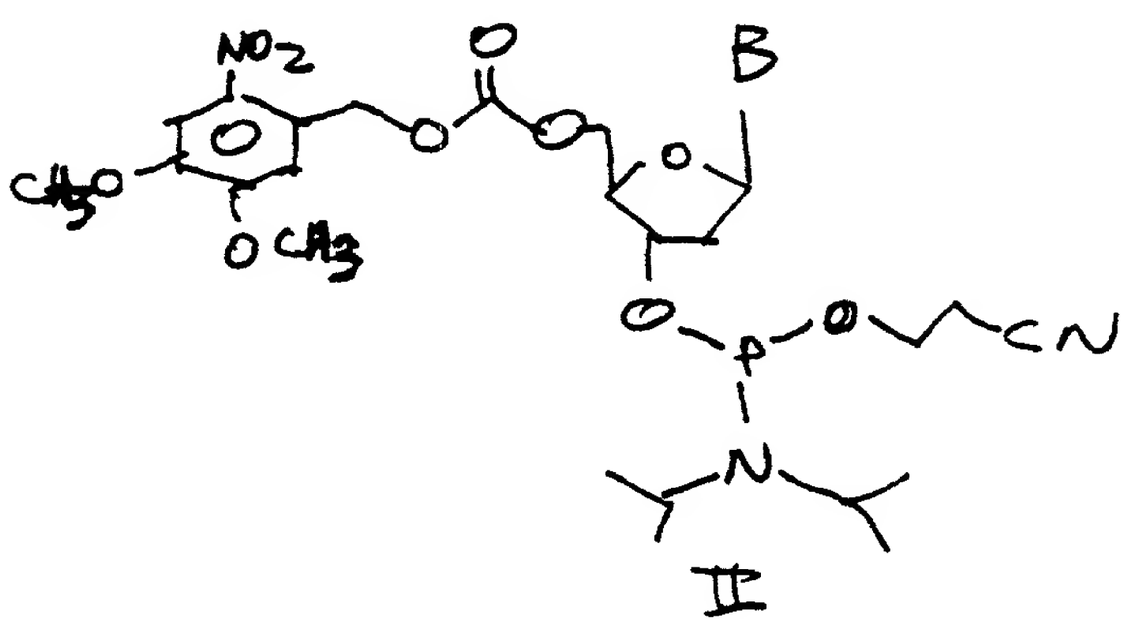
C. Synthetic Methods

The synthetic methods in making a substrate are described in the parent application, U.S.S.N. 07/492,462. The construction of the matrix pattern on the substrate will typically be generated by the use of photo-sensitive reagents. By use of photo-lithographic optical methods, particular segments of the substrate can be irradiated with light to activate or deactivate blocking agents, e.g., to protect or deprotect particular chemical groups. By an appropriate sequence of photo-exposure steps at appropriate times with appropriate masks and with appropriate reagents, the substrates can have known polymers synthesized at positionally defined regions on the substrate. Methods for synthesizing various substrates are described in U.S.S.N. 07/492,462 (VLSIPS CIP) and U.S.S.N. __/___,___, attorney docket number 11509-28 (automated VLSIPS). By a sequential series of these photo-exposure and reaction manipulations, a defined matrix pattern of known sequences may be generated, and is typically referred to as a VLSIPS substrate. In the nucleic acid synthesis

embodiment, nucleosides used in the synthesis of DNA by photolytic methods will typically be one of the two forms shown below:



I

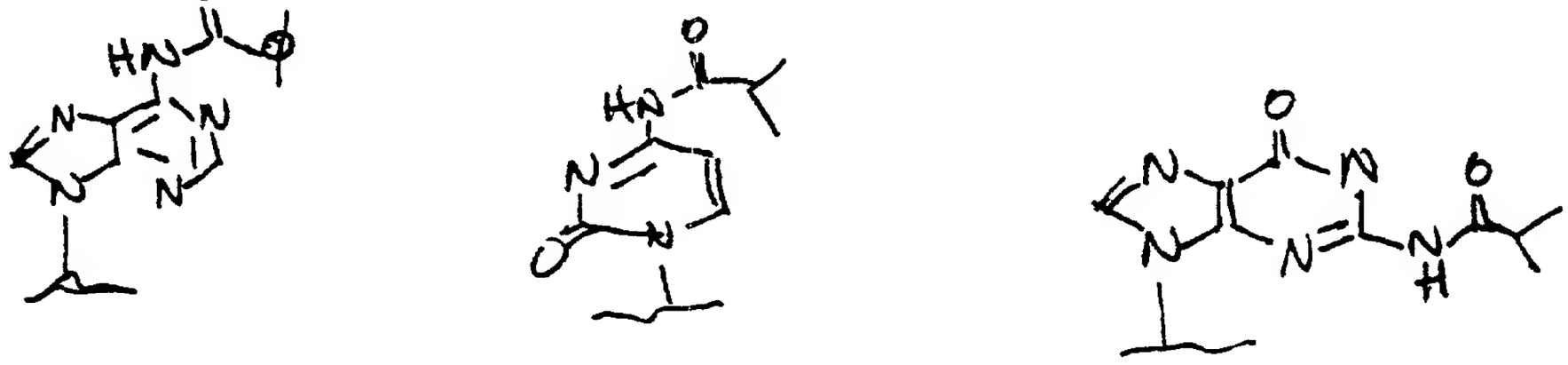


II

B = Adenine, Cytosine, Guanine, or Thymine

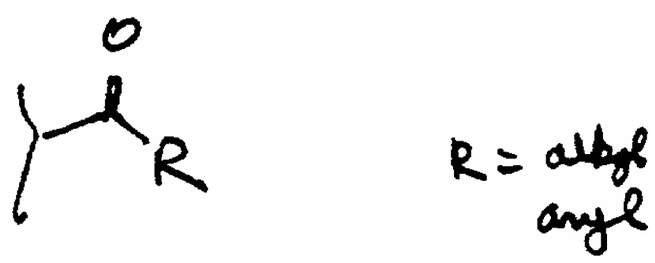
In I, the photolabile group at the 5' position is abbreviated NV (nitroveratryl) and in II, the group is

abbreviated NVOC (nitroveratryl oxycarbonyl). Although not shown in Fig. C the bases (adenine, cytosine, and guanine) contain exocyclic NH_2 groups which must be protected during DNA synthesis. Thymine contains no exocyclic NH_2 and therefore requires no protection. The standard protecting groups for these anaines are shown below:



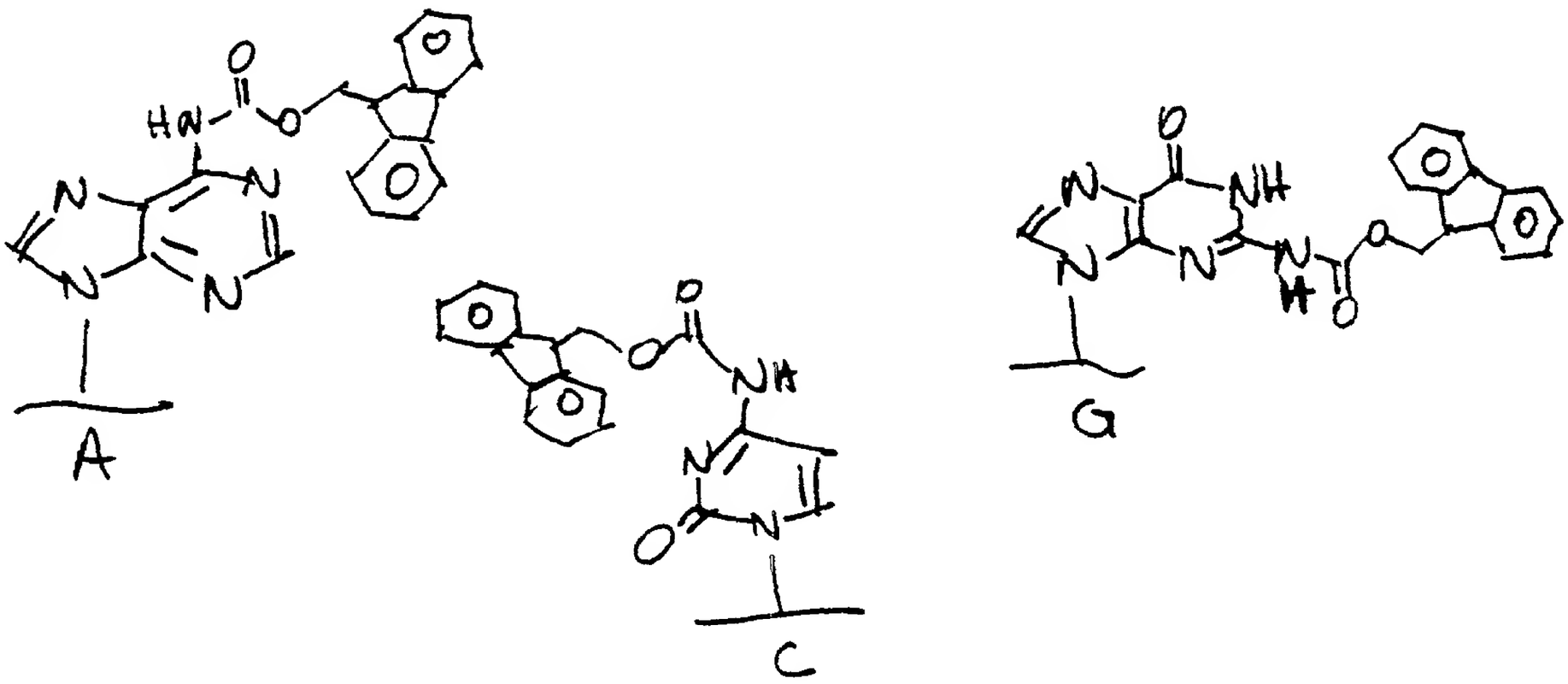
Adenine (A) Cytosine (C) Guanine (G)

Other amides of the general formula



where R may be alkyl or aryl have been used.

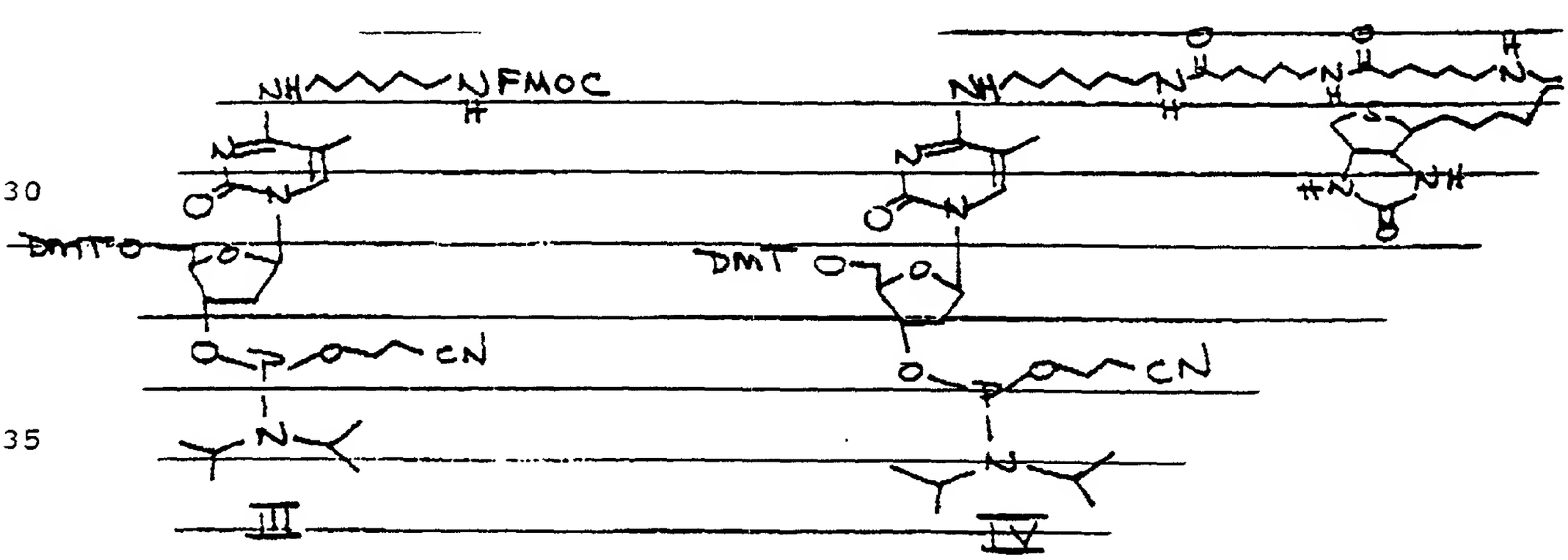
Another type of protecting group FMOC (9-fluorenyl methoxycarbonyl) is currently being used to protect the exocyclic amines of the three bases:



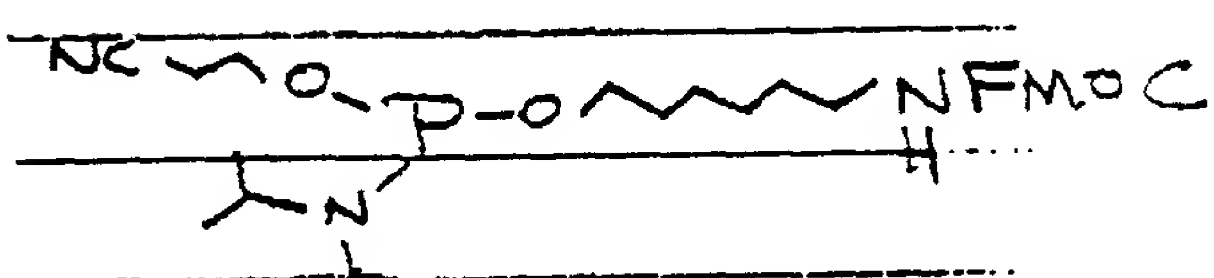
Adenine (A) Cytosine (C) Guanine (G)

The advantage of the FMOC group is that it is removed under mild conditions (dilute organic bases) and can be used for all three bases. The amide protecting groups require more harsh conditions to be removed (NH_3/MeOH with heat).

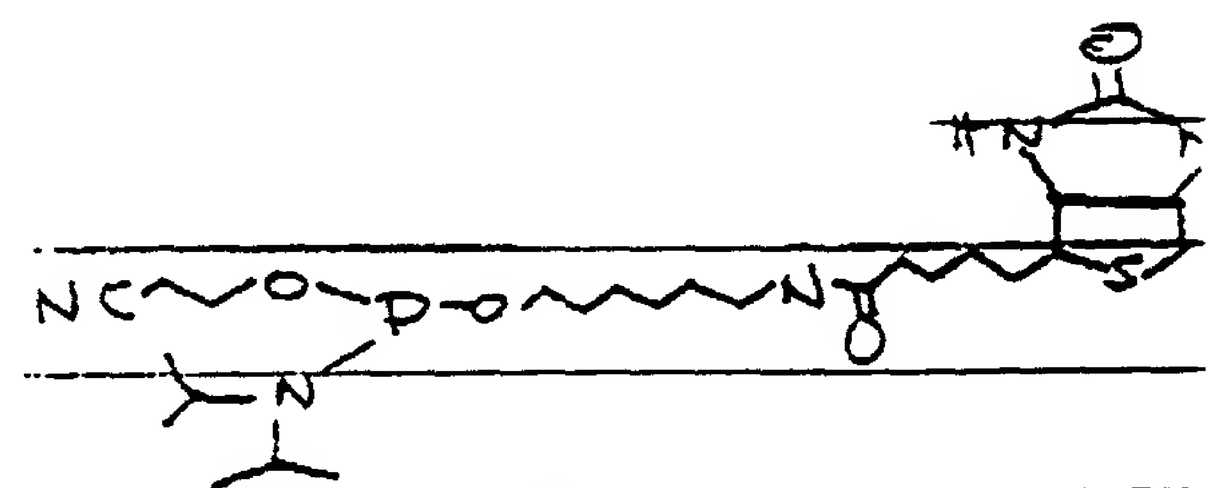
Nucleosides used as 5'-OH probes, useful in verifying correct VLSIPS synthetic function, have been the following:



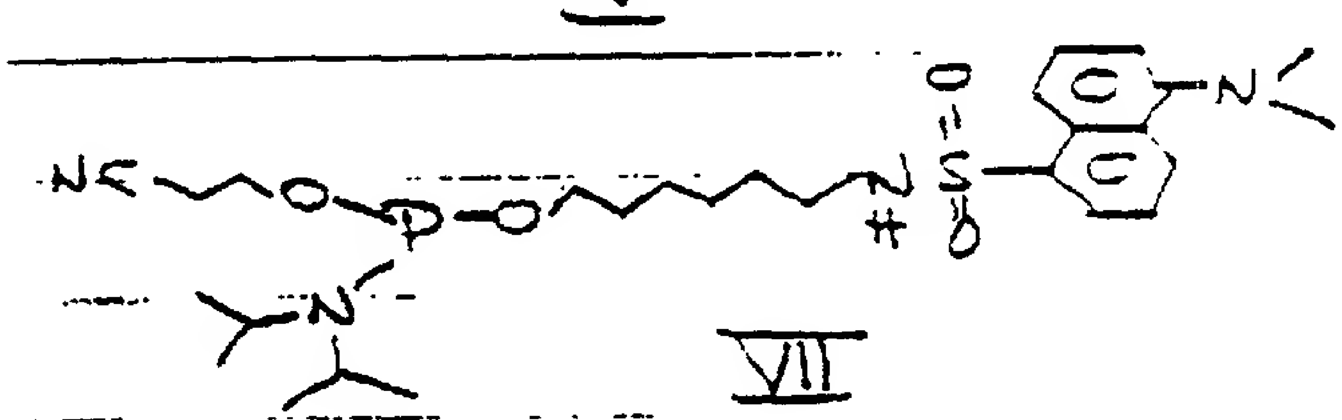
These compounds are used to detect where on a substrate photolysis has occurred by the attachment of either III or V to the newly generated 5'-OH. In the case of III, after the phosphate attachment is made, the substrate is treated with a dilute base to remove the FMOC group. The resulting amine can be reacted with FITC and the substrate examined by fluorescence microscopy. This indicates the proper generation of a 5'-OH. In the case of compound IV, after the phosphate attachment is made, the substrate is treated with FITC labeled streptavidin and the substrate again may be examined by fluorescence microscopy. Other probes, although not nucleoside based, have included the following:



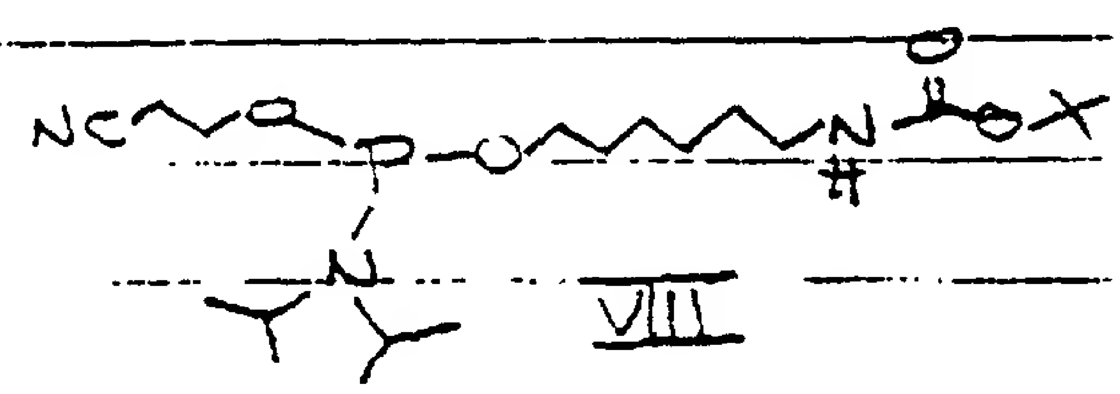
V



VI

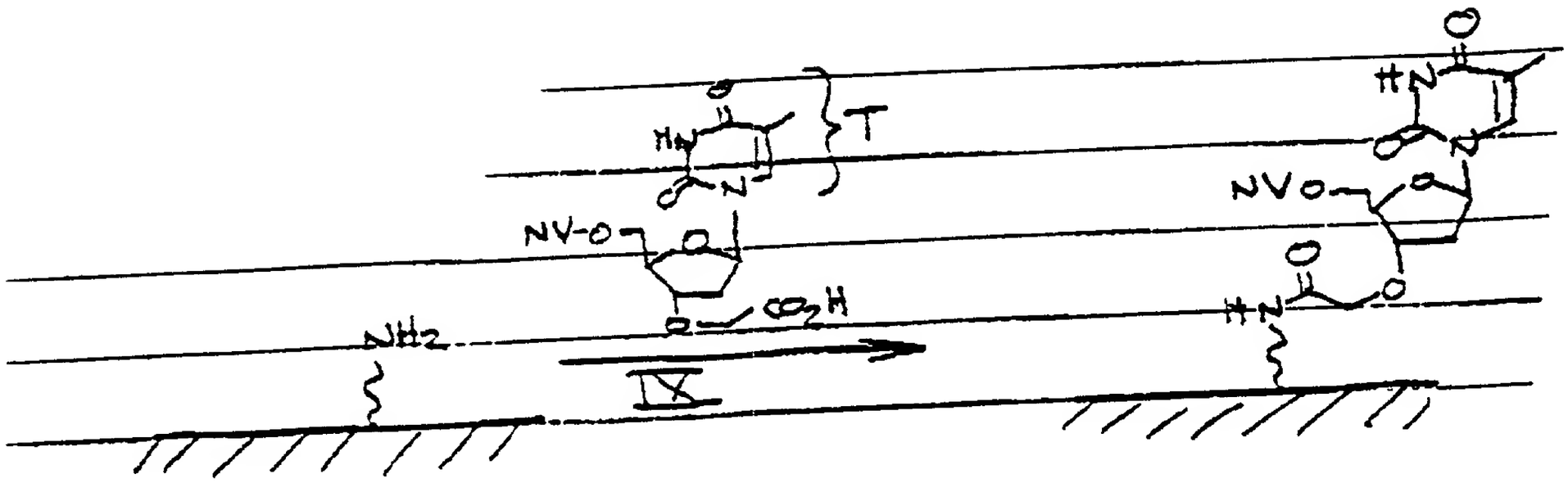


VII

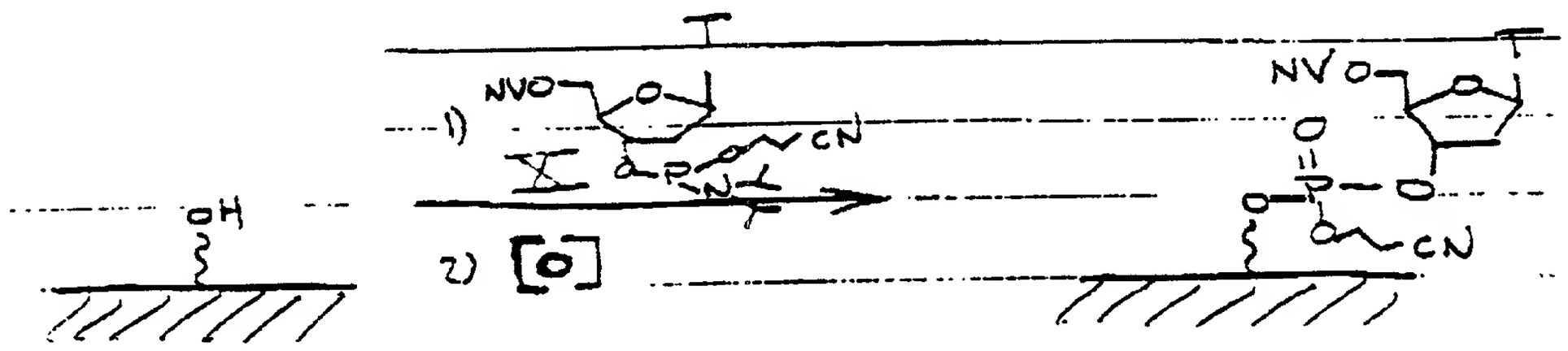


VIII

The method of attachment of the first nucleoside to the surface of the substrate depends on the functionality of the groups at the substrate surface. If the surface is amine functionalized, an amide bond is made (see example below).



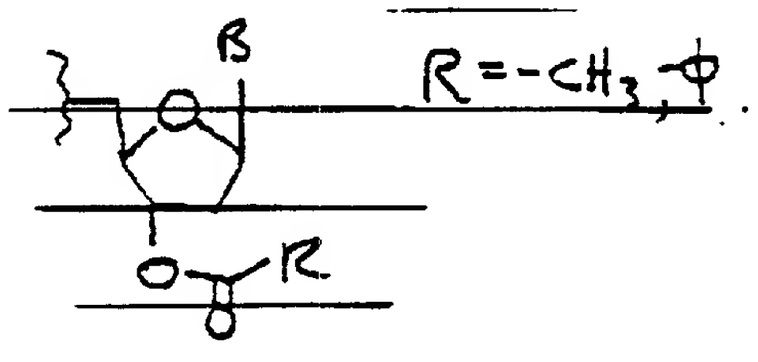
If the surface is hydroxy functionalized a phosphate bond is made (see example below)



In both cases, the thymidine example is illustrated, but any one of the four phosphoramidite activated nucleosides can be used in the first step.

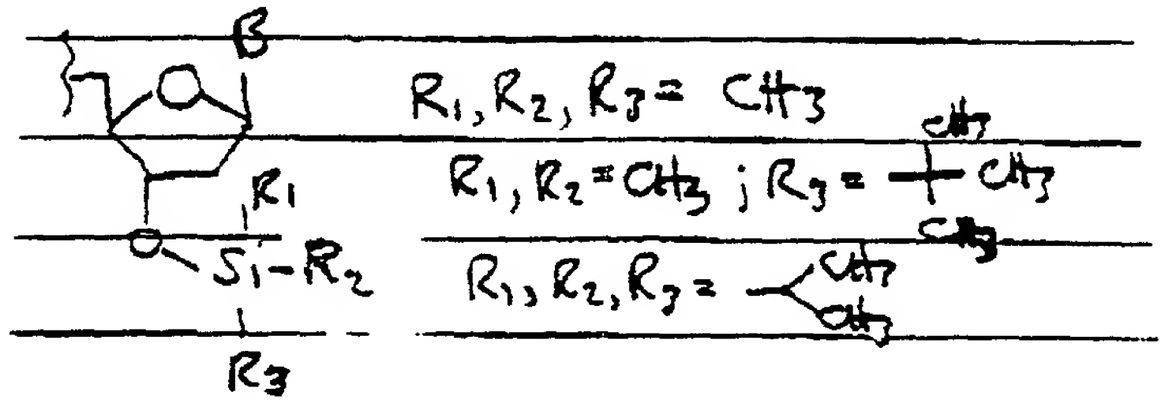
Photolysis of the photolabile group NV or NVOC on the 5' positions of the nucleosides is carried out at ~362 nm with an intensity of 14 mW/cm^2 for 10 minutes with the substrate side (side containing the photolabile group) immersed in dioxane. After the coupling of the next nucleoside is complete, the photolysis is repeated followed by another coupling until the desired oligomer is obtained.

One of the most common 3'-O-protecting group is the ester, in particular the acetate



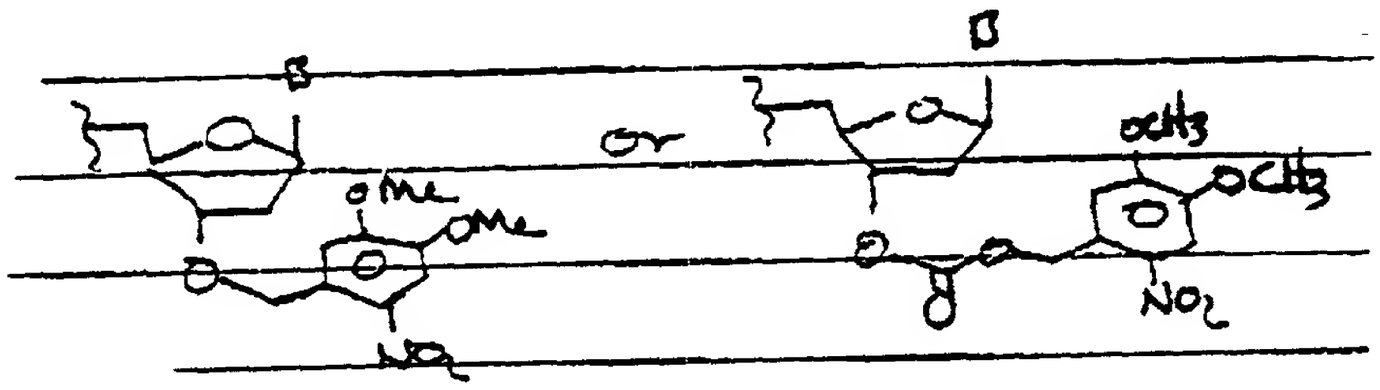
The groups can be removed by mild base treatment 0.1N NaOH/MeOH or $K_2CO_3/H_2O/MeOH$.

Another group used most often is the silyl ether.



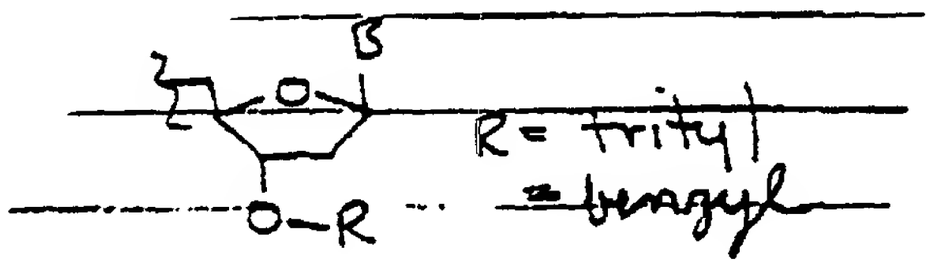
These groups can be removed by neutral conditions using 1 M tetra-n-butylammonium fluoride in THF or under acid conditions.

Related to photodeprotection, the nitroveratryl group could also be used to protect the 3'-position.



Here, light (photolysis) would be used to remove these protecting groups.

A variety of ethers can also be used in the protection of the 3'-O-position.



Removal of these groups usually involves acid or catalytic methods.

Note that corresponding linkages and photoblocked amino acids are described in detail in U.S.S.N. __/___,___, attorney docket number 11509-28, which is hereby incorporated herein by reference.

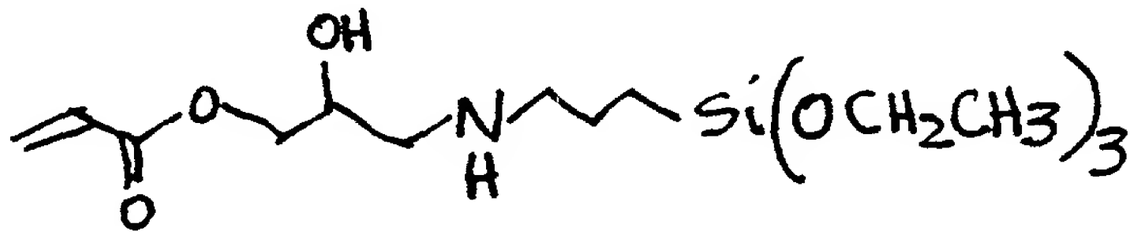
Although the specificity of interactions at particular locations will usually be homogeneous due to a homogeneous polymer being synthesized at each defined location, for certain purposes, it may be useful to have mixed polymers with a commensurate mixed collection of interactions occurring at specific defined locations, or degeneracy reducing analogues, which have been discussed above and show broad specificity in binding. Then, a positive interaction signal may result from any of a number of sequences contained therein.

As an alternative method of generating a matrix pattern on a substrate, preformed polymers may be individually attached at particular sites on the substrate. This may be performed by individually attaching reagents one at a time to specific positions on the matrix, a process which may be automated. See, e.g., U.S.S.N. 07/435,316 (caged biotin parent), and U.S.S.N. 07/612,671 (caged biotin CIP). Another way of generating a positionally defined matrix pattern on a substrate is to have individually specific reagents which interact with each specific position on the substrate. For example, oligonucleotides may be synthesized at defined locations on the substrate. Then the substrate would have on its surface a plurality of regions having homogeneous oligonucleotides attached at each position.

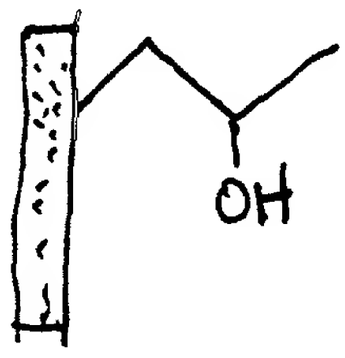
In particular, at least four different substrate preparation procedures are available for treating a substrate surface. They are the standard VLSIPS method, polymeric substrates, Durapore^{TM}, and synthetic beads or fibers. The treatment labeled "standard VLSIPS" method is described in U.S.S.N. __/___,___, attorney docket number 11509-28 (automated VLSIPS), and involves applying amino-propyltriethoxysilane to a glass surface.

The polymeric substrate approach involves either of two ways of generating a polymeric substrate. The first uses a high concentration of aminopropyltriethoxysilane (2-20%) in an aqueous ethanol solution (95%). This allows the silane compound to polymerize both in solution and on the substrate surface, which provides a high density of amines on the surface of the glass. This density is contrasted with the standard VLSIPS method. This polymeric method allows for the deposition on the substrate surface of a monolayer due to the anhydrous method used with the aforementioned silane.

The second polymeric method involves either the coating or covalent binding of an appropriate acrylic acid polymer onto the substrate surface. In particular, e.g., in DNA synthesis, a monomer such as a hydroxypropylacrylate is used to generate a high density of hydroxyl groups on the substrate surface, allowing for the formation of phosphate bonds. An example of such a compound is shown:



The method using a Durapore™ membrane (Millipore) consists of a polyvinylidine difluoride coating with crosslinked polyhydroxylpropyl acrylate [PVDF-HPA]:



Here the building up of, e.g., a DNA oligomer, can be started immediately since phosphate bonds to the surface can be accomplished in the first step with no need for modification.

A nucleotide dimer (5'-C-T-3') has been successfully made on this substrate in our labs.

The fourth method utilizes synthetic beads or fibers. This would use another substrate, such as a teflon copolymer graft bead or fiber, which is covalently coated with an organic layer (hydrophilic) terminating in hydroxyl sites (commercially available from Molecular Brosystems, Inc.) This would offer the same advantage as the Durapore™ membrane, allowing for immediate phosphate linkages, but would give additional contour by the 3-dimensional growth of oligomers.

A matrix pattern of new reagents may be targeted to each specific oligonucleotide position by attaching a complementary oligonucleotide to which the substrate bound form is complementary. For instance, a number of regions may have homogeneous oligonucleotides synthesized at various locations. Oligonucleotide sequences complementary to each of these can be individually generated and linked to a particular specific reagents. Often these specific reagents will be antibodies. As each of these is specific for finding its complementary oligonucleotide, each of the specific reagents will bind through the oligonucleotide to the appropriate matrix position. A single step having a combination of different specific reagents being attached specifically to a particular oligonucleotide will thereby bind to its complement at the defined matrix position. The oligonucleotides will typically then be covalently attached, using, e.g., an acridine dye, for photocrosslinking. Psoralen is a commonly used acridine dye for photocrosslinking purposes, see, e.g., Song et al. (1979) Photochem. Photobiol. 29:1177-1197; Cimino et al. (1985) Ann. Rev. Biochem. 54:1151-1193; Parsons (1980) Photochem. Photobiol. 32:813-821; and Dattagupta et al. (1985) U.S. Pat. No. 4,542,102, and (1987) U.S. Pat. No. 4,713,326; each of which is hereby incorporated herein by reference. This method allows a single attachment manipulation to attach all of the specific reagents to the matrix at defined positions and results in the specific reagents being homogeneously located at defined positions. In many embodiments, the specific reagents will be antibodies.

In an alternative embodiment, antibody molecules may be used to specifically direct binding to defined positions on a substrate. The VLSIPS technology may be used to generate specific epitopes at each position on the substrate. Antibody molecules having specificity of interaction may be used to attach oligonucleotides, thereby avoiding the interference of internal polynucleotide sequences from binding to the substrate complementary oligonucleotides. In fact, the specificity of interaction for positional targeting may be achieved by use of nucleotide analogues which do not interact with the natural nucleotides. For example, other synthetic nucleotides have been made which undergo base pairing, thereby providing the specificity of targeting, but the synthetic nucleotides also do not interact with the natural biological nucleotides. Thus, synthetic oligonucleotides would be useful for attachment to biological nucleotides and specific targeting. Moreover, the VLSIPS synthetic processes would be useful in generating the VLSIPS substrate, and standard oligonucleotide synthesis could be applied, with minor modifications, to produce the complementary sequences which would be attached to other specific reagents.

D. <u>Surface Immobilization</u>

1. caged biotin

An alternative method of attaching reagents in a positionally defined matrix pattern is to use a caged biotin system. See U.S.S.N. 07/612,671 (caged biotin CIP), which is hereby incorporated herein by reference, for additional details on the chemistry and application of caged biotin embodiments. In short, the caged biotin has a photosensitive blocking moiety which prevents the combination of avidin to biotin. At positions where the photo-lithographic process has removed the blocking group, high affinity biotin sites are generated. Thus, by a sequential series of photolithographic deblocking steps interspersed with exposure of those regions to appropriate biotin containing reagents, only those locations where the deblocking takes place will form an avidin-biotin

interaction. Because the avidin-biotin binding is very tight, this will usually be virtually irreversible binding.

2. crosslinked interactions

The surface immobilization may also take place by photo crosslinking of defined oligonucleotides linked to specific reagents. After hybridization of the complementary oligonucleotides, the oligonucleotides may be crosslinked by a reagent by psoralen or another similar type of acridine dye. Other useful cross linking reagents are described in Dattagupta et al. (1985) U.S. Pat. No. 4,542,102, and (1987) U.S. Pat. No. 4,713,326.

In another embodiment, colony or phage plaque transfer of biological polymers may be transferred directly onto a silicon substrate. For example, a colony plate may be transferred onto a substrate having a generic oligonucleotide sequence which hybridizes to another generic complementary sequence contained on all of the vectors into which inserts are cloned. This will specifically only bind those molecules which are actually contained in the vectors containing the desired complementary sequence. This immobilization allows for producing a matrix onto which a sequence specific reagent can bind, or for other purposes. In a further embodiment, a plurality of different vectors each having a specific oligonucleotide attached to the vector may be specifically attached to particular regions on a matrix having a complementary oligonucleotide attached thereto.

VIII. HYBRIDIZATION/SPECIFIC INTERACTION

A. General

As discussed previously in the VLSIPS parent applications, the VLSIPS substrates may be used for screening for specific interactions with sequence specific targets or probes.

In addition, the availability of substrates having the entire repertoire of possible sequences of a defined length opens up the possibility of sequencing by hybridization. This sequence may be de novo determination of an unknown sequence,

particularly of nucleic acid, verification of a sequence determined by another method, or an investigation of changes in a previously sequenced gene, locating and identifying specific changes. For example, often Maxam and Gilbert sequencing techniques are applied to sequences which have been determined by Sanger and Coulson. Each of those sequencing technologies have problems with resolving particular types of sequences. Sequencing by hybridization may serve as a third and independent method for verifying other sequencing techniques. See, e.g., (1988) Science 242:1245.

In addition, the ability to provide a large repertoire of particular sequences allows use of short subsequence and hybridization as a means to fingerprint a sample. This may be used in a nucleic acid, as well as other polymer embodiments. For example, fingerprinting to a high degree of specificity of sequence matching may be used for identifying highly similar samples, e.g., those exhibiting high homology to the selected probes. This may provide a means for determining classifications of particular sequences. This should allow determination of whether particular genomes of bacteria, phage, or even higher cells might be related to one another.

In addition, fingerprinting may be used to identify an individual source of biological sample. See, e.g., Lander, E. (1989) Nature, 339:501-505, and references therein. For example, a DNA fingerprint may be used to determine whether a genetic sample arose from another individual. This would be particularly useful in various sorts of forensic tests to determine, e.g., paternity or sources of blood samples. Significant detail on the particulars of genetic fingerprinting for identification purposes are described in, e.g., Morris et al. (1989) "Biostatistical evolution of evidence from continuous allele frequency distribution DNA probes in reference to disputed paternity of identity," J. Forensic Science 34:1311-1317; and Neufeld et al. (1990) Scientific American 262:46-53; each of which is hereby incorporated herein by reference.

In another embodiment, a fingerprinting-like procedure may be used for classifying cell types by analyzing a pattern of specific nucleic acids present in the cell. A series of antibodies may be used to identify cell markers, e.g., proteins, usually on the cell surface, but intracellular markers may also be used. Antigens which are extracellularly expressed are preferred so cell lysis is unnecessary in the screening, but intracellular markers may also be useful. The markers will usually be proteins, but may be nucleic acids, lipids, metabolites, carbohydrates, or other cellular components. See, e.g., Winkelgren, I. (1990) Science News 136:234-237, which indicates extracellular DNA may common, and suggesting that such might be characteristic of cell types, stage, or physiology. This may also be useful in defining the temporal stage of development of cells, e.g., stem cells or other cells which undergo temporal changes in development. For example, the stage of a cell, or group of cells, may be tested or defined by isolating a sample of mRNA from the population and testing to see what sequences are present in messenger populations. Direct samples, or amplified samples, may be used. Where particular mRNA or other nucleic acid sequences may be characteristic of or shown to be characteristic of particular developmental stages, physiological states, or other conditions, this fingerprinting method may define them. Similar sorts of fingerprinting may be used for determining T-cell classes or perhaps even to generate classification schemes for such proteins as major histocompatibility complex antigens. Thus, the ability to make these substrates allows both the generation of reagents which will be used for defining subclasses or classes of cells or other biological materials, but also provides the mechanisms for selecting those cells which may be found in defined population groups.

Cell classification defined by such a combination of properties, typically expression of extracellular antigens, the present invention also provides the means for isolating homogeneous population of cells. Once the antigenic determinants which define a cell class have been identified, these antigens may be used in a sequential selection process to

isolate only those cells which exhibit the combination of defining structural properties.

The present invention may also be used for mapping sequences within a larger segment. This may be performed by at least two methods, particularly in reference to nucleic acids. Often, enormous segments of DNA are subcloned into a large plurality of subsequences. Ordering these subsequences may be important in determining the overlaps of sequences upon nucleotide determinations. Mapping may be performed by immobilizing particularly large segments onto a matrix using the VLSIPS technology. Alternatively, sequences may be ordered by virtue of subsequences shared by overlapping segments. See, e.g., Craig et al. (1990) Nuc. Acids Res. 18:2653-2660; Michiels et al. (1987) CABIOS 3:203-210; and Olson et al. (1986) Proc. Natl. Acad. Sci. USA 83:7826-7830.

B. Important Parameters

The extent of specific interaction between reagents immobilized to the VLSIPS substrate and another sequence specific reagent may be modified by the conditions of the interaction. Sequencing embodiments typically require high fidelity hybridization and the ability to discriminate perfect matching from imperfect matching. Fingerprinting and mapping embodiments may be performed using less stringent conditions, depending upon the circumstances.

For example, the specificity of antibody/antigen interaction may depend upon such parameters as pH, salt concentration, ionic composition, solvent composition, detergent composition and concentration, and chaotropic agent concentration. See, e.g., Harlow and Lane (1988) Antibodies: A Laboratory Manual, Cold Spring Harbor Press, New York. By careful control of these parameters, the affinity of binding may be mapped across different sequences.

In a nucleic acid hybridization embodiment, the specificity and kinetics of hybridization have been described in detail by, e.g., Wetmur and Davidson (1968) J. Mol. Biol., 31:349-370, Britten and Kohne (1968) Science 161:529-530, and Kanehisa, (1984) Nuc. Acids Res. 12:203-213, each of which is

hereby incorporated herein by reference. Parameters which are well known to affect specificity and kinetics of reaction include salt conditions, ionic composition of the solvent, hybridization temperature, length of oligonucleotide matching sequences, guanine and cytosine (GC) content, presence of hybridization accelerators, pH, specific bases found in the matching sequences, solvent conditions, and addition of organic solvents.

In particular, the salt conditions required for driving highly mismatched sequences to completion typically include a high salt concentration. The typical salt used is sodium chloride (NaCl), however, other ionic salts may be utilized, e.g., KCl. Depending on the desired stringency hybridization, the salt concentration will often be less than about 3 molar, more often less than 2.5 molar, usually less than about 2 molar, and more usually less than about 1.5 molar. For applications directed towards higher stringency matching, the salt concentrations would typically be lower. Ordinary high stringency conditions will utilize salt concentration of less than about 1 molar, more often less then about 750 millimolar, usually less than about 500 millimolar, and may be as low as about 250 or 150 millimolar.

The kinetics of hybridization and the stringency of hybridization both depend upon the temperature at which the hybridization is performed and the temperature at which the washing steps are performed. Temperatures at which steps for low stringency hybridization are desired would typically be lower temperatures, e.g., ordinarily at least about 15°C, more ordinarily at least about 20°C, usually at least about 25°C, and more usually at least about 30°C. For those applications requiring high stringency hybridization, or fidelity of hybridization and sequence matching, temperatures at which hybridization and washing steps are performed would typically be high. For example, temperatures in excess of about 35°C would often be used, more often in excess of about 40°C, usually at least about 45°C, and occasionally even temperatures as high as about 50°C or 60°C or more. Of course, the hybridization of oligonucleotides may be disrupted by even

higher temperatures. Thus, for stripping of targets from substrates, as discussed below, temperatures as high as 80°C, or even higher may be used.

The base composition of the specific oligonucleotides involved in hybridization affects the temperature of melting, and the stability of hybridization as discussed in the above references. However, the bias of GC rich sequences to hybridize faster and retain stability at higher temperatures can be compensated for by the inclusion in the hybridization incubation or wash steps of various buffers. Sample buffers which accomplish this result include the triethly-and trimethyl ammonium buffers. See, e.g., Wood et al. (1987) Proc. Natl. Acad. Sci. USA, 82:1585-1588, and Khrapko, K. et al. (1989) FEBS Letters 256:118-122.

The rate of hybridization can also be affected by the inclusion of particular hybridization accelerators. These hybridization accelerators include the volume exclusion agents characterized by dextran sulfate, or polyethylene glycol (PEG). Dextran sulfate is typically included at a concentration of between 1% and 40% by weight. The actual concentration selected depends upon the application, but typically a faster hybridization is desired in which the concentration is optimized for the system in question. Dextran sulfate is often included at a concentration of between 0.5% and 2% by weight or dextran sulfate at a concentration between about 0.5% and 5%. Alternatively, proteins which accelerate hybridization may be added, e.g., the recA protein found in E. coli) or other homologous proteins.

With respect to those embodiments where specific reagents are not oligonucleotides, the conditions of specific interaction would depend on the affinity of binding between the specific reagent and its target. Typically parameters which would be of particular importance would be pH, salt concentration anion and cation compositions, buffer concentration, organic solvent inclusion, detergent concentration, and inclusion of such reagents such as chaotropic agents. In particular, the affinity of binding may be tested over a variety of conditions by multiple washes and

repeat scans or by using reagents with differences in binding affinity to determine which reagents bind or do not bind under the selected binding and washing conditions. The spectrum of binding affinities may provide an additional dimension of information which may be very useful in identification purposes and mapping.

Of course, the specific hybridization conditions will be selected to correspond to a discriminatory condition which provides a positive signal where desired but fails to show a positive signal at affinities where interaction is not desired. This may be determined by a number of titration steps or with a number of controls which will be run during the hybridization and/or washing steps to determine at what point the hybridization conditions have reached the stage of desired specificity.

IX. DETECTION METHODS

Methods for detection depend upon the label selected. The criteria for selecting an appropriate label are discussed below, however, a fluorescent label is preferred because of its extreme sensitivity and simplicity. Standard labeling procedures are used to determine the positions where interactions between a sequence and a reagent take place. For example, if a target sequence is labeled and exposed to a matrix of different probes, only those locations where probes do interact with the target will exhibit any signal. Alternatively, other methods may be used to scan the matrix to determine where interaction takes place. Of course, the spectrum of interactions may be determined in a temporal manner by repeated scans of interactions which occur at each of a multiplicity of conditions. However, instead of testing each individual interaction separately, a multiplicity of sequence interactions may be simultaneously determined on a matrix.

A. Labeling Techniques

The target polynucleotide may be labeled by any of a number of convenient detectable markers. A fluorescent label is preferred because it provides a very strong signal with low

background. It is also optically detectable at high resolution and sensitivity through a quick scanning procedure. Other potential labeling moieties include, radioisotopes, chemiluminescent compounds, labeled binding proteins, heavy metal atoms, spectroscopic markers, magnetic labels, and linked enzymes.

Another method for labeling may bypass any label of the target sequence. The target may be exposed to the probes, and a double strand hybrid is formed at those positions only. Addition of a double strand specific reagent will detect where hybridization takes place. An intercalative dye such as ethidium bromide may be used as long as the probes themselves do not fold back on themselves to a significant extent forming hairpin loops. See, e.g., Sheldon et al. (1986) U.S. Pat. No. 4,582,789. However, the length of the hairpin loops in short oligonucleotide probes would typically be insufficient to form a stable duplex.

In another embodiment, different targets may be simultaneously sequenced where each target has a different label. For instance, one target could have a green fluorescent label and a second target could have a red fluorescent label. The scanning step will distinguish sites of binding of the red label from those binding the green fluorescent label. Each sequence can be analyzed independently from one another.

Suitable chromogens will include molecules and compounds which absorb light in a distinctive range of wavelengths so that a color may be observed, or emit light when irradiated with radiation of a particular wave length or wave length range, e.g., fluorescers. Biliproteins, e.g., ficoerythrin, may also serve as labels.

A wide variety of suitable dyes are available, being primary chosen to provide an intense color with minimal absorption by their surroundings. Illustrative dye types include quinoline dyes, triarylmethane dyes, acridine dyes, alizarine dyes, phthaleins, insect dyes, azo dyes, anthraquinoid dyes, cyanine dyes, phenazathionium dyes, and phenazoxonium dyes.

A wide variety of fluorescers may be employed either by themselves or in conjunction with quencher molecules. Fluorescers of interest fall into a variety of categories having certain primary functionalities. These primary functionalities include 1- and 2-aminonaphthalene, p,p'-diaminostilbenes, pyrenes, quaternary phenanthridine salts, 9-aminoacridines, p,p'-diaminobenzophenone imines, anthracenes, oxacarbocyanine, merocyanine, 3-aminoequilenin, perylene, bis-benzoxazole, bis-p-oxazolyl benzene, 1,2-benzophenazin, retinol, bis-3-aminopyridinium salts, hellebrigenin, tetracycline, sterophenol, benzimidzaolylphenylamine, 2-oxo-3-chromen, indole, xanthen, 7-hydroxycoumarin, phenoxazine, salicylate, strophanthidin, porphyrins, triarylmethanes and flavin. Individual fluorescent compounds which have functionalities for linking or which can be modified to incorporate such functionalities include, e.g., dansyl chloride; fluoresceins such as 3,6-dihydroxy-9-phenylxanthhydrol; rhodamineisothiocyanate; N-phenyl 1-amino-8-sulfonatonaphthalene; N-phenyl 2-amino-6-sulfonatonaphthalene; 4-acetamido-4-isothiocyanato-stilbene-2,2'-disulfonic acid; pyrene-3-sulfonic acid; 2-toluidinonaphthalene-6-sulfonate; N-phenyl, N-methyl 2-aminoaphthalene-6-sulfonate; ethidium bromide; stebrine; auromine-0,2-(9'-anthroyl)palmitate; dansyl phosphatidylethanolamine; N,N'-dioctadecyl oxacarbocyanine; N,N'-dihexyl oxacarbocyanine; merocyanine, 4-(3'pyrenyl)butyrate; d-3-aminodesoxy-equilenin; 12-(9'-anthroyl)stearate; 2-methylanthracene; 9-vinylanthracene; 2,2'-(vinylene-p-phenylene)bisbenzoxazole; p-bis[2-(4-methyl-5-phenyl-oxazolyl)]benzene; 6-dimethylamino-1,2-benzophenazin; retinol; bis(3'-aminopyridinium) 1,10-decandiyl diiodide; sulfonaphthylhydrazone of hellibrienin; chlorotetracycline; N-(7-dimethylamino-4-methyl-2-oxo-3-chromenyl)maleimide; N-[p-(2-benzimidazolyl)-phenyl]maleimide; N-(4-fluoranthyl)maleimide; bis(homovanillic acid); resazarin; 4-chloro-7-nitro-2,1,3-benzooxadiazole; merocyanine 540; resorufin; rose bengal; and 2,4-diphenyl-3(2H)-furanone.

Desirably, fluorescers should absorb light above about 300 nm, preferably about 350 nm, and more preferably above about 400 nm, usually emitting at wavelengths greater than about 10 nm higher than the wavelength of the light absorbed. It should be noted that the absorption and emission characteristics of the bound dye may differ from the unbound dye. Therefore, when referring to the various wavelength ranges and characteristics of the dyes, it is intended to indicate the dyes as employed and not the dye which is unconjugated and characterized in an arbitrary solvent.

Fluorescers are generally preferred because by irradiating a fluorescer with light, one can obtain a plurality of emissions. Thus, a single label can provide for a plurality of measurable events.

Detectable signal may also be provided by chemiluminescent and bioluminescent sources. Chemiluminescent sources include a compound which becomes electronically excited by a chemical reaction and may then emit light which serves as the detectible signal or donates energy to a fluorescent acceptor. A diverse number of families of compounds have been found to provide chemiluminescence under a variety of conditions. One family of compounds is 2,3-dihydro-1,-4-phthalazinedione. The most popular compound is luminol, which is the 5-amino compound. Other members of the family include the 5-amino-6,7,8-trimethoxy- and the dimethylamino[ca]benz analog. These compounds can be made to luminesce with alkaline hydrogen peroxide or calcium hypochlorite and base. Another family of compounds is the 2,4,5-triphenylimidazoles, with lophine as the common name for the parent product. Chemiluminescent analogs include para-dimethylamino and -methoxy substituents. Chemiluminescence may also be obtained with oxalates, usually oxalyl active esters, e.g., p-nitrophenyl and a peroxide, e.g., hydrogen peroxide, under basic conditions. Alternatively, luciferins may be used in conjunction with luciferase or lucigenins to provide bioluminescence.

Spin labels are provided by reporter molecules with an unpaired electron spin which can be detected by electron

spin resonance (ESR) spectroscopy. Exemplary spin labels include organic free radicals, transitional metal complexes, particularly vanadium, copper, iron, and manganese, and the like. Exemplary spin labels include nitroxide free radicals.

B. Scanning System

With the automated detection apparatus, the correlation of specific positional labeling is converted to the presence on the target of sequences for which the reagents have specificity of interaction. Thus, the positional information is directly converted to a database indicating what sequence interactions have occurred. For example, in a nucleic acid hybridization application, the sequences which have interacted between the substrate matrix and the target molecule can be directly listed from the positional information. The detection system used is described in U.S.S.N. 07/649,642 (VLSIPS CIP); and U.S.S.N. __/___,___, attorney docket number 11509-28 (automated VLSIPS). Although the detection described therein is a fluorescence detector, the detector may be replaced by a spectroscopic or other detector. The scanning system may make use of a moving detector relative to a fixed substrate, a fixed detector with a moving substrate, or a combination. Alternatively, mirrors or other apparatus can be used to transfer the signal directly to the detector. See, e.g, U.S.S.N. __/___,___, attorney docket number 11509-28 (automated VLSIPS), which is hereby incorporated herein by reference.

The detection method will typically also incorporate some signal processing to determine whether the signal at a particular matrix position is a true positive or may be a spurious signal. For example, a signal from a region which has actual positive signal may tend to spread over and provide a positive signal in an adjacent region which actually should not have one. This may occur, e.g., where the scanning system is not properly discriminating with sufficiently high resolution in its pixel density to separate the two regions. Thus, the signal over the spatial region may be evaluated pixel by pixel to determine the locations and the actual extent of positive signal. A true positive signal should, in theory, show a

uniform signal at each pixel location. Thus, processing by plotting number of pixels with actual signal intensity should have a clearly uniform signal intensity. Regions where the signal intensities show a fairly wide dispersion, may be particularly suspect and the scanning system may be programmed to more carefully scan those positions.

In another embodiment, as the sequence of a target is determined at a particular location, the overlap for the sequence would necessarily have a known sequence. Thus, the system can compare the possibilities for the next adjacent position and look at these in comparison with each other. Typically, only one of the possible adjacent sequences should give a positive signal and the system might be programmed to compare each of these possibilities and select that one which gives a strong positive. In this way, the system can also simultaneously provide some means of measuring the reliability of the determination by indicating what the average signal to background ratio actually is.

More sophisticated signal processing techniques can be applied to the initial determination of whether a positive signal exists or not. See, e.g., U.S.S.N. __/___,___, attorney docket number 11509-28 (automated VLSIPS).

From a listing of those sequences which interact, data analysis may be performed on a series of sequences. For example, in a nucleic acid sequence application, each of the sequences may be analyzed for their overlap regions and the original target sequence may be reconstructed from the collection of specific subsequences obtained therein. Other sorts of analyses for different applications may also be performed, and because the scanning system directly interfaces with a computer the information need not be transferred manually. This provides for the ability to handle large amounts of data with very little human intervention. This, of course, provides significant advantages over manual manipulations. Increased throughput and reproducibility is thereby provided by the automation of vast majority of steps in any of these applications.

XI. DATA ANALYSIS

A. General

Data analysis will typically involve aligning the proper sequences with their overlaps to determine the target sequence. Although the target "sequence" may not specifically correspond to any specific molecule, especially where the target sequence is broken and fragmented up in the sequencing process, the sequence corresponds to a contiguous sequence of the subfragments.

The data analysis can be performed by a computer using an appropriate program. See, e.g., Drmanac, R. et al. (1989) Genomics 4:114-128; and a commercially available analysis program available from the Genetic Engineering Center, P.O. Box 794, 11000 Belgrade, Yugoslavia. Although the specific manipulations necessary to reassemble the target sequence from fragments may take many forms, one embodiment uses a sorting program to sort all of the subsequences using a defined hierarchy. The hierarchy need not necessarily correspond to any physical hierarchy, but provides a means to determine, in order, which subfragments have actually been found in the target sequence. In this manner, overlaps can be checked and found directly rather than having to search throughout the entire set after each selection process. For example, where the oligonucleotide probes are 10-mers, the first 9 positions can be sorted. A particular subsequence can be selected as in the examples, to determine where the process starts. As analogous to the theoretical example provided above, the sorting procedure provides the ability to immediately find the position of the subsequence which contains the first 9 positions and can compare whether there exists more than 1 subsequence during the first 9 positions. In fact, the computer can easily generate all of the possible target sequences which contain given combination of subsequences. Typically there will be only one, but in various situations, there will be more.

An exemplary flow chart for a sequencing program is provided in Figure 4. In general terms, the program provides for automated scanning of the substrate to determine the

positions of probe and target interaction. Simple processing of the intensity of the signal may be incorporated to filter out clearly spurious signals. The positions with positive interaction are correlated with the sequence specificity of specific matrix positions, to generate the set of matching subsequences. This information is further correlated with other target sequence information, e.g., restriction fragment analysis. The sequences are then aligned using overlap data, thereby leading to possible corresponding target sequences which will, optimally, correspond to a single target sequence.

B. Hardware

A variety of computer systems may be used to run a sequencing program. The program may be written to provide both the detecting and scanning steps together and will typically be dedicated to a particular scanning apparatus. However, the components and functional steps may be separated and the scanning system may provide an output, e.g., through tape or an electronic connection into a separate computer which separately runs the sequencing analysis program. The computer may be any of a number of machines provided by standard computer manufacturers, e.g., IBM compatible machines, Apple™ machines, VAX machines, and others, which may often use a UNIX™ operating system. Of course, the hardware used to run the analysis program will typically determine what programming language would be used.

C. Software

Software would be easily developed by a person of ordinary skill in the programming art, following the flow chart provided, or based upon the input provided and the desired result.

Of course, an exemplary embodiment is a polynucleotide sequence system. However, the theoretical and mathematical manipulations necessary for data analysis of other linear molecules, such as polypeptides, carbohydrates, and various other polymers are conceptually similar. Simple branching polymers will usually also be sequencable using

similar technology. However, where there is branching, it may be desired that additional recognition reagents be used to determine the nature and location of branches. This can easily be provided by use of appropriate specific reagents which would be generated by methods similar to those used to produce specific reagents for linear polymers.

XII. SUBSTRATE REUSE

Where a substrate is made with specific reagents that are relatively insensitive to the handling and processing steps involved in a single cycle of use, the substrate may often be reused. The target molecules are usually stripped off of the solid phase specific recognition molecules. Of course, it is preferred that the manipulations and conditions be selected as to be mild and to not affect the substrate. For example, if a substrate is acid labile, a neutral pH would be preferred in all handling steps. Similar sensitivities would be carefully respected where recycling is desired.

A. Removal of Label

Typically for a recycling, the previously attached specific interaction would be disrupted and removed. This will typically involve exposing the substrate to conditions under which the interaction between probe and target is disrupted. Alternatively, it may be exposed to conditions where the target is destroyed. For example, where the probes are oligonucleotides and the target is a polynucleotide, a heating and low salt wash will often be sufficient to disrupt the interactions. Additional reagents may be added such as detergents, and organic or inorganic solvents which disrupt the interaction between the specific reagents and target. In an embodiment where the specific reagents are antibodies, the substrate may be exposed to a gentle detergent which will denature the specific binding between the antibody and its target. The conditions are selected to avoid severe disruption or destruction of the structure of the antibody and to maintain the specificity of the antibody binding site. Conditions with specific pH, detergent concentration, salt concentration, ionic

concentration, and other parameters may be selected which disrupt the specific interactions.

B. Storage and Preservation

As indicated above, the matrix will typically be maintained under conditions where the matrix itself and the linkages and specific reagents are preserved. Various specific preservatives may be added which prevent degradation. For example, if the reagents are acid or base labile, a neutral pH buffer will typically be added. It is also desired to avoid destruction of the matrix by growth of organisms which may destroy organic reagents attached thereto. For this reason, a preservative such as cyanide or azide may be added. However, the chemical preservative should also be selected to preserve the chemical nature of the linkages and other components of the substrate. Typically, a detergent may also be included.

C. Processes to Avoid Degradation of Oligomers

In particular, a substrate comprising a large number of oligomers will be treated in a fashion which is known to maintain the quality and integrity of oligonucleotides. These include storing the substrate in a carefully controlled environment under conditions of lower temperature, cation depletion (EDTA and EGTA), sterile conditions, and inert argon or nitrogen atmosphere.

XIII. INTEGRATED SEQUENCING STRATEGY

A. Initial Mapping Strategy

As indicated above, although the VLSIPS may be applied to sequencing embodiments, it is often useful to integrate other concepts to simply the sequencing. For example, nucleic acids may be easily sequenced by careful selection of the vectors and hosts used for amplifying and generating the specific target sequences. For example, it may be desired to use specific vectors which have been designed to interact most efficiently with the VLSIPS substrate. This is also important in fingerprinting and mapping strategies. For example, vectors may be carefully selected having particular

complementary sequences which are designed to attach to a genetic or specific oligomer on the substrate. This is also applicable to situations where it is desired to target particular sequences to specific locations on the matrix.

In one embodiment, unnatural oligomers may be used to target natural probes to specific locations on the VLSIPS substrate. In addition, particular probes may be generated for the mapping embodiment which are designed to have specific combinations of characteristics. For example, the construction of a mapping substrate may depend upon use of another automated apparatus which takes clones isolated from a chromosome walk and attaches them individually or in bulk to the VLSIPS substrate.

In another embodiment, a variety of specific vectors having known and particular "targeting" sequences adjacent the cloning sites may be individually used to clone a selected probe, and the isolated probe will then be targetable to a site on the VLSIPS substrate with a sequence complementary to the "target" sequence.

B. <u>Selection of Smaller Clones</u>

In the fingerprinting and mapping embodiments, the selection of probes may be very important. Significant mathematical analysis may be applied to determine which specific sequences should be used as those probes. Of course, for fingerprinting use, these sequences would be most desired that show significant heterogeneity across the human population. Selection of the specific sequences which would most favorably be utilized will tend to be single copy sequences within the genome.

Various hybridization selection procedures may be applied to select sequences which tend not to be repeated within a genome, and thus would tend to be conserved across individuals. For example, hybridization selections may be made for non-repetitive and single copy sequences. See, e.g., Britten and Kohne (1968) "Repeated Sequences in DNA," <u>Science</u> 161:529-540. On the other hand, it may be desired under certain circumstances to use repeated sequences. For example,

where a fingerprint may be used to identify or distinguish different species, or where repetitive sequences may be diagnostic of specific species, repetitive sequences may be desired for inclusion in the fingerprinting probes. In either case, the sequencing capability will greatly assist in the selection of appropriate sequences to be used as probes.

Also as indicated above, various means for constructing an appropriate substrate may involve either mechanical or automated procedures. The standard VLSIPS automated procedure involves synthesizing oligonucleotides or short polymers directly on the substrate. In various other embodiments, it is possible to attach separately synthesized reagents onto the matrix in an ordered array. Other circumstances may lend themselves to transfer a pattern from a petri plate onto a solid substrate. Also, there are methods for site specifically directing collections of reagents to specific locations using unnatural nucleotides or equivalent sorts of targeting molecules.

While a brute force manual transfer process may be utilized sequentially attaching various samples to successive positions, instrumentation for automating such procedures may also be devised. The automated system for performing such would preferably be relatively easily designed and conceptually easily understood.

XIV. COMMERCIAL APPLICATIONS

A. Sequencing

As indicated above, sequencing may be performed either de novo or as a verification of another sequencing method. The present hybridization technology provides the ability to sequence nucleic acids and polynucleotides de novo, or as a means to verify either the Maxam and Gilbert chemical sequencing technique or Sanger and Coulson dideoxy- sequencing techniques. The hybridization method is useful to verify sequencing determined by any other sequencing technique and to closely compare two similar sequences, e.g., to identify and locate sequence differences.

Besides polynucleotide sequencing, the present invention also provides means for sequencing other polymers. This includes polypeptides, carbohydrates, synthetic organic polymers, and other polymers. Again, the sequencing may be either verification or de novo.

Of course, sequencing of can be very important in many different sorts of environments. For example, it will be useful in determining the genetic sequence of particular markers in various individuals. In addition, polymers may be used as markers or for information containing molecules to encode information. For example, a short polynucleotide sequence may be included in large bulk production samples indicating the manufacturer, date, and location of manufacture of a product. For example, various drugs may be encoded with this information with a small number of molecules in a batch. For example, a pill may have somewhere from 10 to 100 to 1,000 or more very short and small molecules encoding this information. When necessary, this information may be decoded from a sample of the material using a polymerase chain reaction (PCR) or other amplification method. This encoding system may be used to provide the origin of large bulky samples without significantly affecting the properties of those samples. For example, chemical samples may also be encoded by this method thereby providing means for identifying the source and manufacturing details of lots. The origin of bulk hydrocarbon samples may be encoded. Production lots of organic compounds such as benzene or plastics may be encoded with a short molecule polymer. Food stuffs may also be encoded using similar marking molecules. Even toxic waste samples can be encoded determining the source or origin. In this way, proper disposal can be traced or more easily enforced.

Similar sorts of encoding may be provided by fingerprinting-type analysis. Whether the resolution is absolute or less so, the concept of coding information on molecules such as nucleic acids, which can be amplified and later decoded, may be a very useful and important application.

This technology also provides the ability to include markers for origins of biological materials. For example, a

patented animal line may be transformed with a particular unnatural sequence which can be traced back to its origin. With a selection of multiple markers, the likelihood could be negligible that a combination of markers would have independently arisen from a source other than the patented or specifically protected source. This technique may provide a means for tracing the actual origin of particular biological materials. Bacteria, plants, and animals will be subject to marking by such encoding sequences.

B. Fingerprinting

As indicated above, fingerprinting technology may also be used for data encryption. Moreover, fingerprinting allows for significant identification of particular individuals. Where the fingerprinting technology is standardized, and used for identification of large numbers of people, related equipment and peripheral processing will be developed to accompany the underlying technology. For example, specific equipment may be developed for automatically taking a biological sample and generating or amplifying the information molecules within the sample to be used in fingerprinting analysis. Moreover, the fingerprinting substrate may be mass produced using particular types of automatic equipment. Synthetic equipment may produce the entire matrix simultaneously by stepwise synthetic methods as provided by the VLSIPS technology. The attachment of specific probes onto a substrate may also be automated, e.g., making use of the caged biotin technology. See, e.g., U.S.S.N. 07/612,671 (caged biotin CIP). As indicated above, there are automated methods for actually generating the matrix and substrate with distinct sequence reagents positionally located at each of the matrix positions. Where such reagents are, e.g., unnatural amino acids, a targeting function may be utilized which does not interfere with aa natural nucleotide functionality.

In addition, peripheral processing may be important and may be dedicated to this specific application. Thus, automated equipment for producing the substrates may be designed, or particular systems which take in a biological

sample and output either a computer readout or an encoded instrument, e.g., a card or document which indicates the information and can provide that information to others. An identification having a short magnetic strip with a few million bits may be used to provide individual identification and important medical information useful in a medical emergency.

In fact, data banks may be set up to correlate all of this information of fingerprinting with medical information. This may allow for the determination of correlations between various medical problems and specific DNA sequences. By collating large populations of medical records with genetic information, genetic propensities and genetic susceptibilities to particular medical conditions may be developed. Moreover, with standardization of substrates, the micro encoding data may be also standardized to reproduce the information from a centralized data bank or on an encoding device carried on an individual person. On the other hand, if the fingerprinting procedure is sufficiently quick and routine, every hospital may routinely perform a fingerprinting operation and from that determine many important medical parameters for an individual.

In particular industries, the VLSIPS sequencing, fingerprinting, or mapping technology will be particularly appropriate. As mentioned above, agricultural livestock suppliers may be able to encode and determine whether their particular strains are being used by others. By incorporating particular markers into their genetic stocks, the markers will indicate origin of genetic material. This is applicable to seed producers, livestock producers, and other suppliers of medical or agricultural biological materials.

This may also be useful in identifying individual animals or plants. For example, these markers may be useful in determining whether certain fish return to their original breeding grounds, whether sea turtles always return to their original birthplaces, or to determine the migration patterns and viability of populations of particular endangered species. It would also provide means for tracking the sources of particular animal products. For example, it might be useful for determining the origins of controlled animal substances

such as elephant ivory or particular bird populations whose importation or exportation is controlled.

As indicated above, polymers may be used to encode important information on source and batch and supplier. This is described in greater detail, e.g., "Applications of PCR to industrial problems," (1990) in Chemical and Engineering News 68:145, which is hereby incorporated herein by reference. In fact, the synthetic method can be applied to the storage of enormous amounts of information. Small substrates may encode enormous amounts of information, and its recovery will make use of the inherent replication capacity. For example, on regions of 10 μm x 10 μm, 1 cm^2 has 10^6 regions. An theory, the entire human genome could be attached in 1000 nucleotide segments on a 3 cm^2 surface. Genomes of endangered species may be stored on these substrates.

Fingerprinting may also be used for genetic tracing or for identifying individuals for forensic science purposes. See, e.g., Morris, J. et al. (1989) "Biostatistical Evaluation of Evidence From Continuous Allele Frequency Distribution DNA Probes in Reference to Disputed Paternity and Identity," J. Forensic Science 34:1311-1317, and references provided therein; each of which is hereby incorporated herein by reference.

In addition, the high resolution fingerprinting allows the distinguishability to high resolution of particular samples. As indicated above, new cell classifications may be defined based on combinations of a large number of properties. Similar applications will be found in distinguishing different species of animals or plants. In fact, microbial identification may become dependent on characterization of the genetic content. Tumors or other cells exhibiting abnormal physiology will be detectable by use of the present invention. Also, knowing the genetic fingerprint of a microorganism may provide very useful information on how to treat an infection by such organism.

Modifications of the fingerprint embodiments may be used to diagnose the condition of the organism. For example, a blood sample is presently used for diagnosing any of a number of different physiological conditions. A multi-dimensional

fingerprinting method made available by the present invention could become a routine means for diagnosing an enormous number of physiological features simultaneously. This may revolutionize the practice of medicine in providing information on an enormous number of parameters together at one time. In another way, the genetic predisposition may also revolutionize the practice of medicine providing a physician with the ability to predict the likelihood of particular medical conditions arising at any particular moment. It also provides the ability to apply preventative medicine.

The present invention might also find application in use for screening new drugs and new reagents which may be very important in medical diagnosis or other applications. For example, a description of generating a population of monoclonal antibodies with defined specificities may be very useful for producing various drugs or diagnostic reagents.

Also available are kits with the reagents useful for performing sequencing, fingerprinting, and mapping procedures. The kits will have various compartments with the desired necessary reagents, e.g., substrate, labeling reagents for target samples, buffers, and other useful accompanying products.

C. Mapping

The present invention also provides the means for mapping sequences within enormous stretches of sequence. For example, nucleotide sequences may be mapped within enormous chromosome size sequence maps. For example, it would be possible to map a chromosomal location within the chromosome which contains hundreds of millions of nucleotide base pairs. In addition, the mapping and fingerprinting embodiments allow for testing of chromosomal translocations, one of the standard problems for which amniocentesis is performed.

Thus, the present invention provides a powerful tool and the means for performing sequencing, fingerprinting, and mapping functions on polymers. Although most easily and directly applicable to polynucleotides, polypeptides,

carbohydrates, and other sorts of molecules can be advantageously utilized using the present technology.

The present invention will be better understood by reference to the following illustrative examples. The following examples are offered by way of illustration and not by way of limitation.

EXPERIMENTAL

I. Sequencing
A. polynucleotide
B. polypeptide
C. short peptide
1. Herz antibody identification
II. Fingerprinting
A. polynucleotide fingerprint
B. peptide fingerprint
C. cell classification scheme
D. temporal development scheme
1. developmental antigens
2. developmental mRNA expression
E. diagnostic test
1. viral identification
2. bacterial identification
3. other microbiological identifications
4. allergy test (immobilized antigens)
F. individual (animal/plant) identification
1. genetic
2. immunological
G. genetic screen
1. test alleles with markers
2. amniocentesis
III. Mapping
A. positionally located clones (caged biotin)
1. short probes, long targets
2. long targets, short probes
B. positionally defined clones
IV. Conclusion

* * * * * * * * * * * * * * * *

Relevant applications whose techniques are incorporated herein by reference are Pirrung, et al., U.S.S.N. 07/362,901 (VLSIPS parent), filed June 7, 1989; Pirrung et al, U.S.S.N. 07/492,462 (VLSIPS CIP), filed March 7, 1990; Barrett, et al., U.S.S.N. 07/435,316 (caged biotin) filed November 13, 1989; Barrett, et al., U.S.S.N. 07/612,671 (caged biotin CIP), filed November 13, 1990; and commonly assigned and simultaneously filed applications U.S.S.N. __/___,___, attorney docket number 11509-28 (automated VLSIPS) and U.S.S.N. __/___,___, attorney docket number 11509-26 (sequencing by synthesis).

Also, additional relevant techniques are described, e.g., in Sambrook, J., et al. (1989) Molecular Cloning: a Laboratory Manual, 2d Ed., vols 1-3, Cold Spring Harbor Press,

New York; Greenstein and Winitz (1961) Chemistry of the Amino Acids, Wiley and Sons, New York; Bodzansky, M. (1988) Peptide Chemistry: a Practical Textbook, Springer-Verlag, New York; Harlow and Lane (1988) Antibodies: A Laboratory Manual, Cold Spring Harbor Press, New York; Glover, D. (ed.) (1987) DNA Cloning: A Practical Approach, vols 1-3, IRL Press, Oxford; Bishop and Rawlings (1987) Nucleic Acid and Protein Sequence Analysis: A Practical Approach, IRL Press, Oxford; Hames and Higgins (1985) Nucleic Acid Hybridisation: A Practical Approach, IRL Press, Oxford; Wu et al. (1989) Recombinant DNA Methodology, Academic Press, San Diego; Goding (1986) Monoclonal Antibodies: Principles and Practice, (2d ed.), Academic Press, San Diego; Finegold and Barron (1986) Bailey and Scott's Diagnostic Microbiology, (7th ed.), Mosby Co., St. Louis; Collins et al. (1989) Microbiological Methods, (6th ed.), Butterworth, London; Chaplin and Kennedy (1986) Carbohydrate Analysis: A Practical Approach, IRL Press, Oxford; Van Dyke (ed.) (1985) Bioluminescence and Chemiluminescence: Instruments and Applications, vol 1, CRC Press, Boca Rotan; and Ausubel et al. (ed.) (1990) Current Protocols in Molecular Biology, Greene Publishing and Wiley-Interscience, New York; each of which is hereby incorporated herein by reference.

The following examples are provided to illustrate the efficacy of the inventions herein. All operations were conducted at about ambient temperatures and pressures unless indicated to the contrary.

I. SEQUENCING

A. Polynucleotide

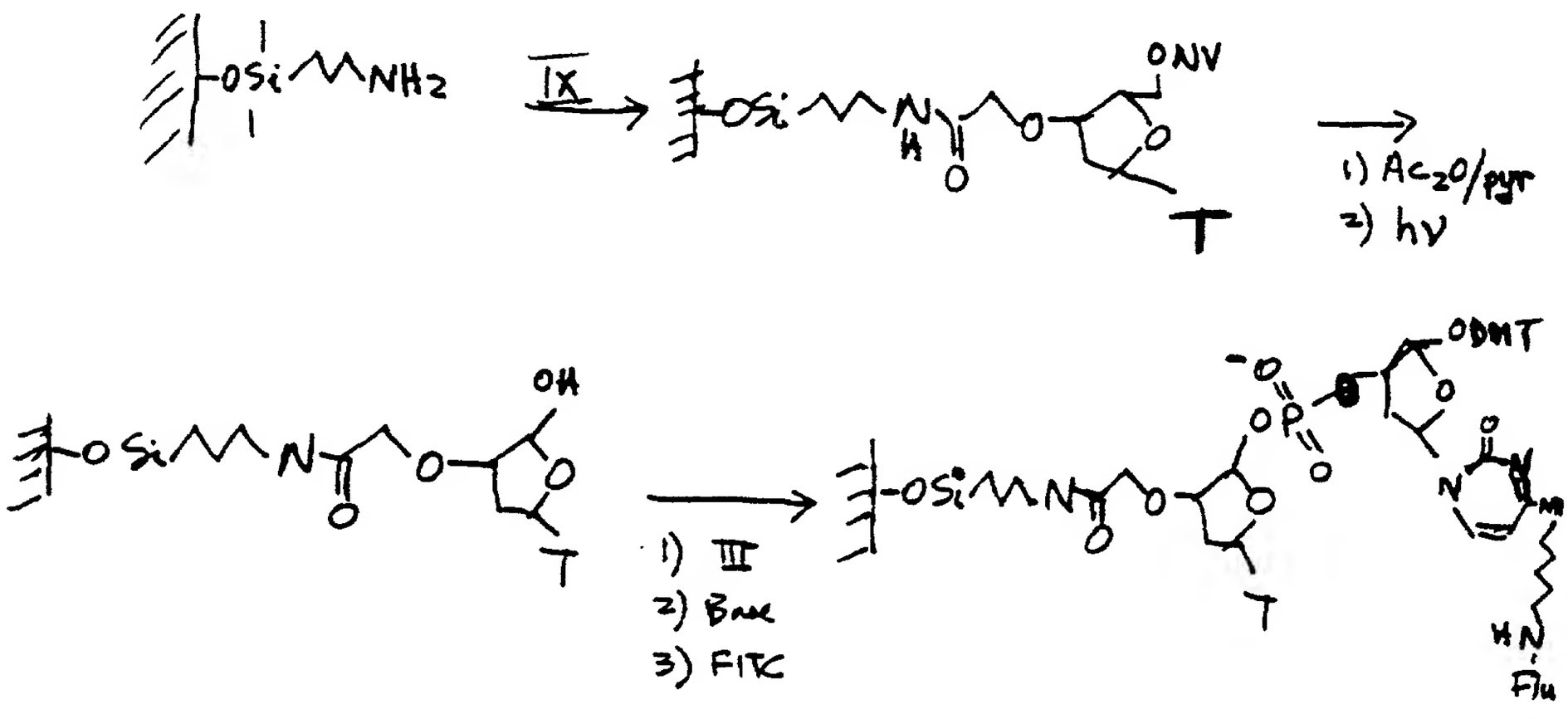
1. HPLC of the photolysis of 5'-O-nitroveratryl-thymidine.

In order to determine the time for photolysis of 5'-O-nitrovertryl thymidine to thymidine a 100 μM solution of NV-Thym-OH (5'-O-nitrovertryl thymidine) in dioxane was made and ~200 μl aliquots were irradiated (in a quartz cuvette 1 cm x 2 mm) at 362.3 nm for 20 sec, 40 sec, 60 sec, 2 min, 5 min, 10 min, 15 min, and 20 min. The resulting irradiated mixtures were then analyzed by HPLC using a Varian MicroPak SP column

(C_{18} analytical) at a flow rate of 1 ml/min and a solvent system of 40% CH_3CN and 60% water. Thymidine has a retention time of 1.2 min and NVO-Thym-OH has a retention time of 2.1 min. It was seen that after 10 min of exposure the deprotection was complete.

2. Preparation and Detection of Thymidine-Cytidine dimer (FITC)

The reaction is illustrated:



To an aminopropylated glass slide (standard VLSIPS) was added a mixture of the following:

12.2 mg of NVO-Thym-CO_2H (IX)
3.4 mg of HOBT (N-hydroxybenztriazal)
8.8 μl DIEA (Diisopropylethylamine)
11.1 mg BOP reagent
2.5 ml DMF

After 2 h coupling time (standard VLSIPS) the plate was washed, acetylated with acetic anhydride/pyridine, washed,

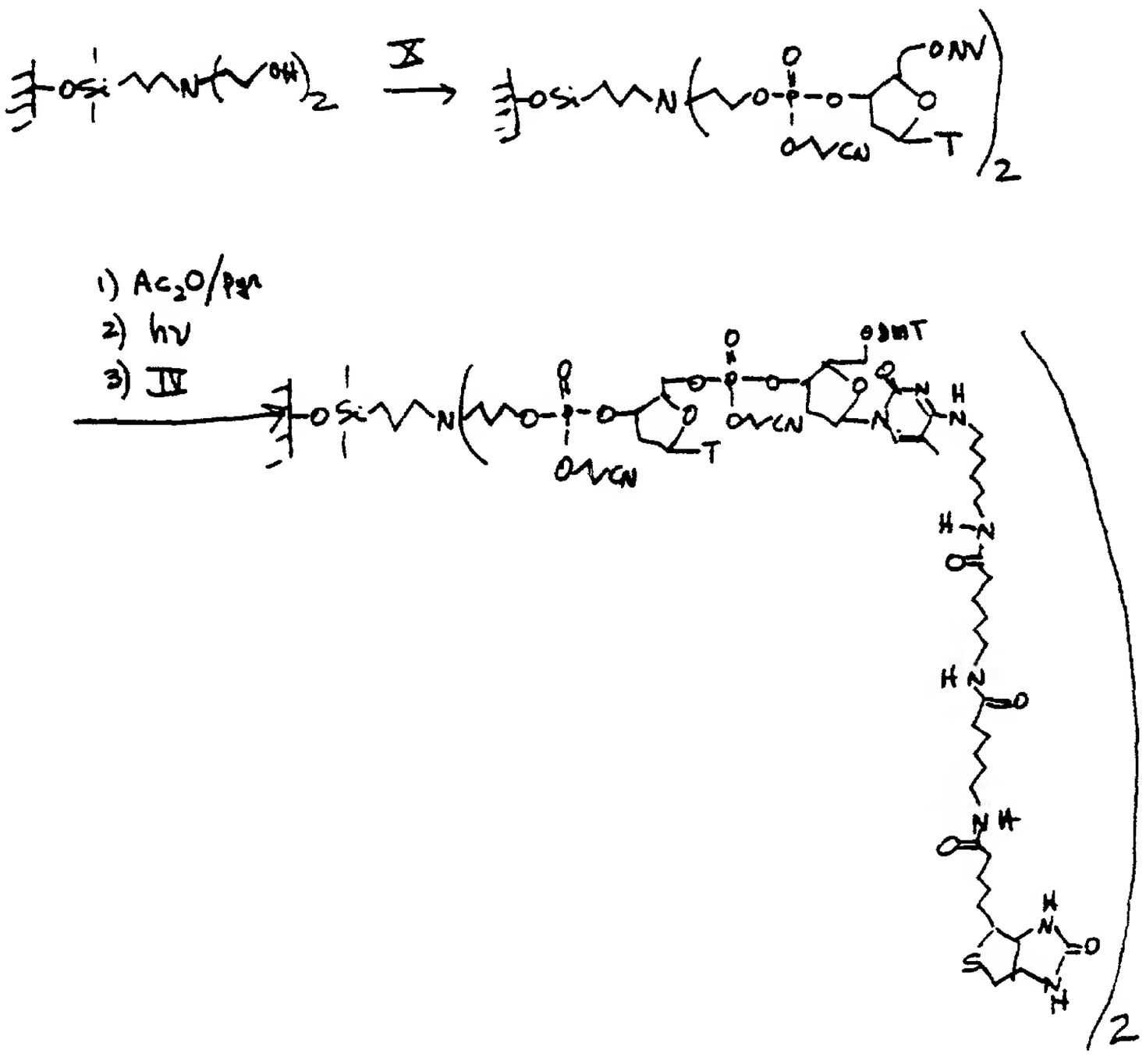
dried, and photolyzed in dioxane at 362 nm at 14 mW/cm^2 for 10 min using a 500 μm checkerboard mask. The slide was then taken and treated with a mixture of the following:

107 mg of FMOC-amine modified C (III)

21 mg of tetrazole

1 ml anhydrous CH_3CN

After being treated for approximately 8 min, the slide was washed off with CH_3CN, dried, and oxidized with I_2/H_2O/THF/lutidine for 1 min. The slide was again washed, dried, and treated for 30 min with a 20% solution of DBU in DMF. After thorough rinsing of the slide, it was next exposed to a FITC solution (1mM fluorescein isothiocyanate [FITC] in DMF) for 50 min, then washed, dried, and examined by fluorescence microscopy. This reaction is illustrated:



3. Preparation and Detection of Thymidine-Cytidine dimer (Biotin)

An aminopropyl glass slide, was soaked in a solution of ethylene oxide (20% in DMF) to generate a hydroxylated surface. The slide was added a mixture of the following:

32 mg of NVO-T-OCED (X)

11 mg of tetrazole

0.5 ml of anhydrous CH_3CN

After 8 min the plate was then rinsed with acetonitrile, then oxidized with I_2/H_2O/THF/lutidine for 1 min, washed and dried. The slide was then exposed to a 1:3 mixture of acetic anhydride:pyridine for 1 h, then washed and dried. The substrate was a then photolyzed in dioxane at 362 nm at 14 mW/cm^2 for 10 min using a 500μm checkerboard mask, dried, and then treated with a mixture of the following:

65 mg of biotin modified C (IV)

11 mg of tetrazole

0.5 ml anhydrous CH_3CN

After 8 min the slide was washed with CH_3CN then oxidized with I_2/H_2O/THF/lutidine for 1 min, washed, and then dried. The slide was then soaked for 30 min in a PBS/0.05% Tween 20 buffer and the solution then shaken off. The slide was next treated with FITC-labeled streptavidin at 10 μg/ml in the same buffer system for 30 min. After this time the streptavidin-buffer system was rinsed off with fresh PBS/0.05% Tween 20 buffer and then the slide was finally agitated in distilled water for about 1/2 h. After drying, the slide was examined by fluorescence microscopy (see Fig. 2 and Fig. 3).

4. substrate preparation

Before attachment of reactive groups it is preferred to clean the substrate which is, in a preferred embodiment, a glass substrate such as a microscope slide or cover slip. A roughened surface will be useable but a plastic or other solid substrate is also appropriate. According to one embodiment the slide is soaked in an alkaline bath consisting of, e.g., 1 liter of 95% ethanol with 120 ml of water and 120 grams of sodium hydroxide for 12 hours. The slides are washed with a

buffer and under running water, allowed to air dry, and rinsed with a solution of 95% ethanol.

The slides are then aminated with, e.g., aminopropyltriethoxysilane for the purpose of attaching amino groups to the glass surface on linker molecules, although other omega functionalized silanes could also be used for this purpose. In one embodiment 0.1% aminopropyltriethoxysilane is utilized, although solutions with concentrations from 10^{-7}% to 10% may be used, with about 10^{-3}% to 2% preferred. A 0.1% mixture is prepared by adding to 100 ml of a 95% ethanol/5% water mixture, 100 microliters (μl) of aminopropyltriethoxysilane. The mixture is agitated at about ambient temperature on a rotary shaker for an appropriate amount of time, e.g., about 5 minutes. 500 μl of this mixture is then applied to the surface of one side of each cleaned slide. After 4 minutes or more, the slides are decanted of this solution and thoroughly rinsed three times or more by dipping in 100% ethanol.

After the slides dry, they are heated in a 110-120°C vacuum oven for about 20 minutes, and then allowed to cure at room temperature for about 12 hours in an argon environment. The slides are then dipped into DMF (dimethylformamide) solution, followed by a thorough washing with methylene chloride.

5. linker attachment, blocking of free sites

The aminated surface of the slide is then exposed to about 500 μl of, for example, a 30 millimolar (mM) solution of NVOC-nucleotide- NHS (N-hydroxysuccinimide) in DMF for attachment of a NVOC-nucleotide to each of the amino groups. See, e.g., SIGMA Chemical Company for various nucleotide derivatives. The surface is washed with, for example, DMF, methylene chloride, and ethanol.

Any unreacted aminopropyl silane on the surface, i.e., those amino groups which have not had the NVOC-nucleotide attached, are now capped with acetyl groups (to prevent further reaction) by exposure to a 1:3 mixture of acetic anhydride in pyridine for 1 hour. Other materials which may perform this

residual capping function include trifluoroacetic anhydride, formicacetic anhydride, or other reactive acylating agents. Finally, the slides are washed again with DMF, methylene chloride, and ethanol.

6. synthesis of eight trimers of C and T

Fig. 4 illustrates a possible synthesis of the eight trimers of the two-monomer set: cytosine and thymine (represented by C and T, respectively). A glass slide bearing silane groups terminating in 6-nitroveratryloxycarboxamide (NVOC-NH) residues is prepared as a substrate. Active esters (pentafluorophenyl, OBt, etc.) of cytosine and thymine protected at the 5' hydroxyl group with NVOC are prepared as reagents. While not pertinent to this example, if side chain protecting groups are required for the monomer set, these must not be photoreactive at the wavelength of light used to protect the primary chain.

For a monomer set of size n, n x ℓ cycles are required to synthesize all possible sequences of length ℓ. A cycle consists of:

1. Irradiation through an appropriate mask to expose the 5'-OH groups at the sites where the next residue is to be added, with appropriate washes to remove the by-products of the deprotection.
2. Addition of a single activated and protected (with the same photochemically-removable group) monomer, which will react only at the sites addressed in step 1, with appropriate washes to remove the excess reagent from the surface.

The above cycle is repeated for each member of the monomer set until each location on the surface has been extended by one residue in one embodiment. In other embodiments, several residues are sequentially added at one location before moving on to the next location. Cycle times will generally be limited by the coupling reaction rate, now as short as about 10 min in automated oligonucleotide synthesizers. This step is optionally followed by addition of

Of course, greater diversity is obtained by using masking strategies which will also include the synthesis of polymers having a length of less than ℓ. If, in the extreme case, all polymers having a length less than or equal to ℓ are synthesized, the number of polymers synthesized will be:

$$n^{\ell} + n^{\ell-1} + \ldots + n^{1}. \quad (3)$$

The maximum number of lithographic steps needed will generally be n for each "layer" of monomers, i.e., the total number of masks (and, therefore, the number of lithographic steps) needed will be n x ℓ. The size of the transparent mask regions will vary in accordance with the area of the substrate available for synthesis and the number of sequences to be formed. In general, the size of the synthesis areas will be:

size of synthesis areas = (A)/(S)

where:

A is the total area available for synthesis; and
S is the number of sequences desired in the area.

It will be appreciated by those of skill in the art that the above method could readily be used to simultaneously produce thousands or millions of oligomers on a substrate using the photolithographic techniques disclosed herein. Consequently, the method results in the ability to practically test large numbers of, for example, di, tri, tetra, penta, hexa, hepta, octa, nona, deca, even dodecanucleotides, or larger polynucleotides (or correspondingly, polypeptides).

The above example has illustrated the method by way of a manual example. It will of course be appreciated that automated or semi-automated methods could be used. The substrate would be mounted in a flow cell for automated addition and removal of reagents, to minimize the volume of reagents needed, and to more carefully control reaction conditions. Successive masks will be applicable manually or automatically. See, e.g., U.S.S.N. 07/492,462 (VLSIPS CIP) and U.S.S.N. __/___,___, attorney docket number 11509-28 (automated VLSIPS).

7. labeling of target

The target oligonucleotide can be labeled using standard procedures referred to above. As discussed, for certain situations, a reagent which recognizes interaction, e.g., ethidium bromide, may be provided in the detection step. Alternatively, fluorescence labeling techniques may be applied, see, e.g., Smith, et al. (1986) Nature, 321: 674-679; and Prober, et al. (1987) Science, 238:336-341. The techniques described therein will be followed with minimal modifications as appropriate for the label selected.

8. dimers of A, C, G, and T

The described technique may be applied, with photosensitive blocked nucleotides corresponding to adenine, cytosine, guanine, and thymine, to make combinations of polynucleotides consisting of each of the four different nucleotides. All 16 possible dimers would be made using a minor modification of the described method.

9. 10-mers of A, C, G, and T

The described technique for making dimers of A, C, G, and T may be further extended to make longer oligonucleotides. The automated system described, e.g., in U.S.S.N 07/492,462 (VLSIPS CIP), and U.S.S.N. __/___,___, attorney docket number 11509-28 (automated VLSIPS), can be adapted to make all possible 10-mers composed of the 4 nucleotides A, C, G, and T. The photosensitive, blocked nucleotide analogues have been described above, and would be readily adaptable to longer oligonucleotides.

10. specific recognition hybridization to 10-mers

The described hybridization conditions are directly applicable to the sequence specific recognition reagents attached to the substrate, produced as described immediately above. The 10-mers have an inherent property of hybridizing to a complementary sequence. For optimum discrimination between full matching and some mismatch, the conditions of hybridization should be carefully selected, as described above. Careful control of the conditions, and titration of parameters

should be performed to determine the optimum collective conditions.

11. hybridization

Hybridization conditions are described in detail, e.g., in Hames and Higgins (1985) Nucleic Acid Hybridisation: A Practical Approach; and the considerations for selecting particular conditions are described, e.g., in Wetmur and Davidson, (1988) J. Mol. Biol. 31:349-370, and Wood et al. (1985) Proc. Natl. Acad. Sci. USA 82:1585-1588. As described above, conditions are desired which can distinguish matching along the entire length of the probe from where there is one or more mismatched bases. The length of incubation and conditions will be similar, in many respects, to the hybridization conditions used in Southern blot transfers. Typically, the GC bias may be minimized by the introduction of appropriate concentrations of the alkylammonium buffers, as described above.

Titration of the temperature and other parameters is desired to determine the optimum conditions for specificity and distinguishability of absolutely matched hybridization from mismatched hybridization.

A fluorescently labeled target or set of targets are generated, as described in Prober, et al. (1987) Science 238:336-341, or Smith, et al. (1986) Nature 321:674-679. Preferably, the target or targets are of the same length as, or slightly longer, than the oligonucleotide probes attached to the substrate and they will have known sequences. Thus, only a few of the probes hybridize perfectly with the target, and which particular ones did would be known.

The substrate and probes are incubated under appropriate conditions for a sufficient period of time to allow hybridization to completion. The time is measured to determine when the probe-target hybridizations have reached completion. A salt buffer which minimizes GC bias is preferred, incorporating, e.g., buffer, such as tetramethyl ammonium or tetraethyl ammonium ion at between about 2.4 and 3.0 M. See Wood, et al. (1985) Proc. Nat'l Acad. Sci. USA 82:1585-1588.

This time is typically at least about 30 min, and may be as long as about 1-5 days. Typically very long matches will hybridize more quickly, very short matches will hybridize less quickly, depending upon relative target and probe concentrations. The hybridization will be performed under conditions where the reagents are stable for that time duration.

Upon maximal hybridization, the conditions for washing are titrated. Three parameters initially titrated are time, temperature, and cation concentration of the wash step. The matrix is scanned at various times to determine the conditions at which the distinguishability between true perfect hybrid and mismatched hybrid is optimized. These conditions will be preferred in the sequencing embodiments.

12. positional detection of specific interaction

As indicated above, the detection of specific interactions may be performed by detecting the positions where the labeled target sequences are attached. Where the label is a fluorescent label, the apparatus described, e.g., in U.S.S.N. 07/492,462 (VLSIPS CIP); and U.S.S.N. __/___,___, attorney docket number 11509-28, may be advantageously applied. In particular, the synthetic processes described above will result in a matrix pattern of specific sequences attached to the substrate, and a known pattern of interactions can be converted to corresponding sequences.

In an alternative embodiment, a separate reagent which differentially interacts with the probe and interacted probe/targets can indicate where interaction occurs or does not occur. A single-strand specific reagent will indicate where no interaction has taken place, while a double-strand specific reagent will indicate where interaction has taken place. An intercalating dye, e.g., ethidium bromide, may be used to indicate the positions of specific interaction.

13. analysis

Conversion of the positional data into sequence specificity will provide the set of subsequences whose analysis

by overlap segments, may be performed, as described above. Analysis is provided by the methodology described above, or using, e.g., software available from the Genetic Engineering Center, P.O. Box 794, 11000 Belgrade, Yugoslavia (Yugoslav group). See, also, Macevicz, PCT publication no. WO 90/04652, which is hereby incorporated herein by reference.

B. Polypeptide

The description of the preparation of short peptides on a substrate incorporates by reference sections in U.S.S.N. 07/492,462 (VLSIPS CIP), and described below.

1. slide preparation

Preparation of the substrate follows that described above for nucleotides.

2. linker attachment, blocking of free sites

The aminated surface of the slide is exposed to about 500 μl of, e.g., a 30 millimolar (mM) solution of NVOC-GABA (gamma amino butyric acid) NHS (N-hydroxysuccinimide) in DMF for attachment of a NVOC-GABA to each of the amino groups. The surface is washed with, for example, DMF, methylene chloride, and ethanol. See U.S.S.N. __/___,___, attorney docket number 11509-28, for details on amino acid chemistry.

Any unreacted aminopropyl silane on the surface, i.e., those amino groups which have not had the NVOC-GABA attached, are now capped with acetyl groups (to prevent further reaction) by exposure to a 1:3 mixture of acetic anhydride in pyridine for 1 hour. Other materials which may perform this residual capping function include trifluoroacetic anhydride, formicacetic anhydride, or other reactive acylating agents. Finally, the slides are washed again with DMF, methylene chloride, and ethanol.

3. synthesis of 8 trimers of "A" and "B"

See U.S.S.N. 07/492,462 (VLSIPS CIP) which describes the preparation of glycine and phenylalanine trimers. The technique is similar to the method described above for making

triners of C and T, but substituting photosensitive blocked glycine for the C derivative and photosensitive blocked phenylalamine for the T derivative.

4. synthesis of a dimer of an aminopropyl group and a fluorescent group

In synthesizing the dimer of an aminopropyl group and a fluorescent group, a functionalized durapore membrane was used as a substrate. The Durapore membrane was a polyvinylidine difluoride with aminopropyl groups. The aminopropyl groups were protected with the DDZ group by reaction of the carbonyl chloride with the amino groups, a reaction readily known to those of skill in the art. The surface bearing these groups was placed in a solution of THF and contacted with a mask bearing a checkerboard pattern of 1 mm opaque and transparent regions. The mask was exposed to ultraviolet light having a wavelength down to at least about 280 nm for about 5 minutes at ambient temperature, although a wide range of exposure times and temperatures may be appropriate in various embodiments of the invention. For example, in one embodiment, an exposure time of between about 1 and 5000 seconds may be used at process temperatures of between -70 and +50°C.

In one preferred embodiment, exposure times of between about 1 and 500 seconds at about ambient pressure are used. In some preferred embodiments, pressure above ambient is used to prevent evaporation.

The surface of the membrane was then washed for about 1 hour with a fluorescent label which included an active ester bound to a chelate of a lanthanide. Wash times will vary over a wide range of values from about a few minutes to a few hours. These materials fluoresce in the red and the green visible region. After the reaction with the active ester in the fluorophore was complete, the locations in which the fluorophore was bound could be visualized by exposing them to ultraviolet light and observing the red and the green fluorescence. It was observed that the derivatized regions of the substrate closely corresponded to the original pattern of the mask.

5. demonstration of signal capability

Signal detection capability was demonstrated using a low-level standard fluorescent bead kit manufactured by Flow Cytometry Standards and having model no. 824. This kit includes 5.8 μm diameter beads, each impregnated with a known number of fluorescein molecules.

One of the beads was placed in the illumination field on the scan stage in a field of a laser spot which was initially shuttered. After being positioned in the illumination field, the photon detection equipment was turned on. The laser beam was unblocked and it interacted with the particle bead, which then fluoresced. Fluorescence curves of beads impregnated with 7,000 and 29,000 fluorescein molecules, are shown in Figs. 11A and 11B, respectively of U.S.S.N. 07/492,462 (VLSIPS CIP). On each curve, traces for beads without fluorescein molecules are also shown. These experiments were performed with 488 nm excitation, with 100 μW of laser power. The light was focused through a 40 power 0.75 NA objective.

The fluorescence intensity in all cases started off at a high value and then decreased exponentially. The fall-off in intensity is due to photobleaching of the fluorescein molecules. The traces of beads without fluorescein molecules are used for background subtraction. The difference in the initial exponential decay between labeled and nonlabeled beads is integrated to give the total number of photon counts, and this number is related to the number of molecules per bead. Therefore, it is possible to deduce the number of photons per fluorescein molecule that can be detected. This calculation indicates the radiation of about 40 to 50 photons per fluorescein molecule are detected.

6. determination of the number of molecules per unit area

Aminopropylated glass microscope slides prepared according to the methods discussed above were utilized in order to establish the density of labeling of the slides. The free amino termini of the slides were reacted with FITC (fluorescein

isothiocyanate) which forms a covalent linkage with the amino group. The slide is then scanned to count the number of fluorescent photons generated in a region which, using the estimated 40-50 photons per fluorescent molecule, enables the calculation of the number of molecules which are on the surface per unit area.

A slide with aminopropyl silane on its surface was immersed in a 1 mM solution of FITC in DMF for 1 hour at about ambient temperature. After reaction, the slide was washed twice with DMF and then washed with ethanol, water, and then ethanol again. It was then dried and stored in the dark until it was ready to be examined.

Through the use of curves similar to those shown in Fig. 11 of U.S.S.N. 07/492,462 (VLSIPS CIP), and by integrating the fluorescent counts under the exponentially decaying signal, the number of free amino groups on the surface after derivitization was determined. It was determined that slides with labeling densities of 1 fluorescein per 10^3x10^3 to ~2x2 nm could be reproducibly made as the concentration of aminopropyltriethoxysilane varied from 10^{-5}% to 10^{-1}%.

7. removal of NOVC and attachment of a fluorescent marker

NVOC-GABA groups were attached as described above. The entire surface of one slide was exposed to light so as to expose a free amino group at the end of the gamma amino butyric acid. This slide, and a duplicate which was not exposed, were then exposed to fluorescein isothiocyanate (FITC).

Fig. 12A of U.S.S.N. 07/492,462 (VLSIPS CIP) illustrates the slide which was not exposed to light, but which was exposed to FITC. The units of the x axis are time and the units of the y axis are counts. The trace contains a certain amount of background fluorescence. The duplicate slide was exposed to 350 nm broadband illumination for about 1 minute (12 mW/cm^2, ~350 nm illumination), washed and reacted with FITC. A large increase in the level of fluorescence is observed, which indicates photolysis has exposed a number of amino groups on the surface of the slides for attachment of a fluorescent marker.

8. use of a mask in removal of NVOC

The next experiment was performed with a 0.1% aminopropylated slide. Light from a Hg-Xe arc lamp was imaged onto the substrate through a laser-ablated chrome-on-glass mask in direct contact with the substrate.

This slide was illuminated for approximately 5 minutes, with 12 mW of 350 nm broadband light and then reacted with the 1 mM FITC solution. It was put on the laser detection scanning stage and a graph was plotted as a two-dimensional representation of position color-coded for fluorescence intensity. The experiment was repeated a number of times through various masks. The fluorescence patterns for a 100x100 μm mask, a 50 μm mask, a 20 μm mask, and a 10 μm mask indicate that the mask pattern is distinct down to at least about 10 μm squares using this lithographic technique.

9. attachment of YGGFL and subsequent exposure to herz antibody and goat anti-mouse antibody

In order to establish that receptors to a particular polypeptide sequence would bind to a surface-bound peptide and be detected, Leu enkephalin was coupled to the surface and recognized by an antibody. A slide was derivatized with 0.1% amino propyl-triethoxysilane and protected with NVOC. A 500 μm checkerboard mask was used to expose the slide in a flow cell using backside contact printing. The Leu enkephalin sequence (H_2N-tyrosine,glycine,glycine,phenylalanine,leucine-COOH, otherwise referred to herein as YGGFL) was attached via its carboxy end to the exposed amino groups on the surface of the slide. The peptide was added in DMF solution with the BOP/HOBT/DIEA coupling reagents and recirculated through the flow cell for 2 hours at room temperature.

A first antibody, known as the Herz antibody, was applied to the surface of the slide for 45 minutes at 2 μg/ml in a supercocktail (containing 1% BSA and 1% ovalbumin also in this case). A second antibody, goat anti-mouse fluorescein conjugate, was then added at 2 μg/ml in the supercocktail buffer, and allowed to incubate for 2 hours.

The results of this experiment were plotted as fluorescence intensity as a function of position. This image was taken at 10 μm steps and showed that not only can deprotection be carried out in a well defined pattern, but also that (1) the method provided for successful coupling of peptides to the surface of the substrate, (2) the surface of a bound peptide was available for binding with an antibody, and (3) that the detection apparatus capabilities were sufficient to detect binding of a receptor. Moreover, the Herz antibody is a sequence specific reagent which may be used advantageously as a sequence specific recognition reagent. It may be used, if specificity is high, for sequencing purposes, and, at least, for fingerprinting and mapping uses.

10. monomer-by-monomer formation of YGGFL and subsequent exposure to labeled antibody

Monomer-by-monomer synthesis of YGGFL and GGFL in alternate squares was performed on a slide in a checkerboard pattern and the resulting slide was exposed to the Herz antibody.

A slide is derivatized with the aminopropyl group, protected in this case with t-BOC (t-butoxycarbonyl). The slide was treated with TFA to remove the t-BOC protecting group. E-aminocaproic acid, which was t-BOC protected at its amino group, was then coupled onto the aminopropyl groups. The aminocaproic acid serves as a spacer between the aminopropyl group and the peptide to be synthesized. The amino end of the spacer was deprotected and coupled to NVOC-leucine. The entire slide was then illuminated with 12 mW of 325 nm broadband illumination. The slide was then coupled with NVOC-phenylalanine and washed. The entire slide was again illuminated, then coupled to NVOC-glycine and washed. The slide was again illuminated and coupled to NVOC-glycine to form the sequence shown in the last portion of Fig. 13A of U.S.S.N. 07/492,462 (VLSIPS CIP).

Alternating regions of the slide were then illuminated using a projection print using a 500x500 μm checkerboard mask; thus, the amino group of glycine was exposed only in the lighted areas. When the next coupling chemistry

step was carried out, NVOC-tyrosine was added, and it coupled only at those spots which had received illumination. The entire slide was then illuminated to remove all the NVOC groups, leaving a checkerboard of YGGFL in the lighted areas and in the other areas, GGFL. The Herz antibody (which recognizes the YGGFL, but not GGFL) was then added, followed by goat anti-mouse fluorescein conjugate.

The resulting fluorescence scan showed dark areas containing the tetrapeptide GGFL, which is not recognized by the Herz antibody (and thus there is no binding of the goat anti-mouse antibody with fluorescein conjugate), and red areas in which YGGFL was present. The YGGFL pentapeptide is recognized by the Herz antibody and, therefore, there is antibody in the lighted regions for the fluorescein-conjugated goat anti-mouse to recognize.

Similar patterns for a 50 μm mask used in direct contact ("proximity print") with the substrate provided a pattern which was more distinct and the corners of the checkerboard pattern were touching as a result of the mask being placed in direct contact with the substrate (which reflects the increase in resolution using this technique).

11. monomer-by-monomer synthesis of YGGFL and PGGFL

A synthesis using a 50 μm checkerboard mask was conducted. However, P was added to the GGFL sites on the substrate through an additional coupling step. P was added by exposing protected GGFL to light through a mask, and subsequence exposure to P in the manner set forth above. Therefore, half of the regions on the substrate contained YGGFL and the remaining half contained PGGFL.

The fluorescence plot for this experiment showed the regions are again readily discernable between those in which binding did and did not occur. This experiment demonstrated that antibodies are able to recognize a specific sequence and that the recognition is not length-dependent.

12. monomer-by-monomer synthesis of YGGFL and YPGGFL

In order to further demonstrate the operability of the invention, a 50 μm checkerboard pattern of alternating YGGFL and YPGGFL was synthesized on a substrate using techniques like those set forth above. The resulting fluorescence plot showed that the antibody was clearly able to recognize the YGGFL sequence and did not bind significantly at the YPGGFL regions.

13. synthesis of an array of sixteen different amino acid sequences and estimation of relative binding affinity to herz antibody

Using techniques similar to those set forth above, an array of 16 different amino acid sequences (replicated four times) was synthesized on each of two glass substrates. The sequences were synthesized by attaching the sequence NVOC-GFL across the entire surface of the slides. Using a series of masks, two layers of amino acids were then selectively applied to the substrate. Each region had dimensions of 0.25 cm x 0.0625 cm. The first slide contained amino acid sequences containing only L- amino acids while the second slide contained selected D- amino acids. Various regions on the first and second slides, were duplicated four times on each slide. The slides were then exposed to the Herz antibody and fluorescein-labeled goat anti-mouse antibodies.

A fluorescence plot of the first slide, which contained only L- amino acids showed red areas (indicating strong binding, i.e., 149,000 counts or more) and black areas (indicating little or no binding of the Herz antibody, i.e., 20,000 counts or less). The sequence YGGFL was clearly most strongly recognized. The sequences YAGFL and YSGFL also exhibited strong recognition of the antibody. By contrast, most of the remaining sequences showed little or no binding. The four duplicate portions of the slide were extremely consistent in the amount of binding shown therein.

A fluorescence plot of the D- amino acid slide indicated that strongest binding was exhibited by the YGGFL sequence. Significant binding was also detected to YaGFL,

YsGFL, and YpGFL. The remaining sequences showed less binding with the antibody. Low binding efficiency of the sequence yGGFL was observed.

Table 6 lists the various sequences tested in order of relative fluorescence, which provides information regarding relative binding affinity.

Table 6

Apparent Binding to Herz Ab

L- a.a. Set	D- a.a. Set
YGGFL	YGGFL
YAGFL	YaGFL
YSGFL	YsGFL
LGGFL	YpGFL
FGGFL	fGGFL
YPGFL	yGGFL
LAGFL	faGFL
FAGFL	wGGFL
WGGFL	yaGFL
	fpGFL
	waGFL

14. illustrative alternative embodiment

According to an alternative embodiment of the invention, the methods provide for attaching to the surface a caged binding member which, in its caged form, has a relatively low affinity for other potentially binding species, such as receptors and specific binding substances. Such techniques are more fully described in copending application Serial No. 404,920, filed September 8, 1989, and incorporated herein by reference for all purposes. See also U.S.S.N. 07/435,316 (caged biotin parent) and U.S.S.N. 07/612,671 (caged biotin CIP), each of which is hereby incorporated herein by reference.

According to this alternative embodiment, the invention provides methods for forming predefined regions on a surface of a solid support, wherein the predefined regions are capable of immobilizing receptors. The methods make use of caged binding members attached to the surface to enable selective activation of the predefined regions. The caged binding members are liberated to act as binding members ultimately capable of binding receptors upon selective activation of the predefined regions. The activated binding members are then used to immobilize specific molecules such as

receptors on the predefined region of the surface. The above procedure is repeated at the same or different sites on the surface so as to provide a surface prepared with a plurality of regions on the surface containing, for example, the same or different receptors. When receptors immobilized in this way have a differential affinity for one or more ligands, screenings and assays for the ligands can be conducted in the regions of the surface containing the receptors.

The alternative embodiment may make use of novel caged binding members attached to the substrate. Caged (unactivated) members have a relatively low affinity for receptors of substances that specifically bind to uncaged binding members when compared with the corresponding affinities of activated binding members. Thus, the binding members are protected from reaction until a suitable source of energy is applied to the regions of the surface desired to be activated. Upon application of a suitable energy source, the caging groups labilize, thereby presenting the activated binding member. A typical energy source will be light.

Once the binding members on the surface are activated they may be attached to a receptor. The receptor chosen may be a monoclonal antibody, a nucleic acid sequence, a drug receptor, etc. The receptor will usually, though not always, be prepared so as to permit attaching it, directly or indirectly, to a binding member. For example, a specific binding substance having a strong binding affinity for the binding member and a strong affinity for the receptor or a conjugate of the receptor may be used to act as a bridge between binding members and receptors if desired. The method uses a receptor prepared such that the receptor retains its activity toward a particular ligand.

Preferably, the caged binding member attached to the solid substrate will be a photoactivatable biotin complex, i.e., a biotin molecule that has been chemically modified with photoactivatable protecting groups so that it has a significantly reduced binding affinity for avidin or avidin analogs than does natural biotin. In a preferred embodiment, the protecting groups localized in a predefined region of the

surface will be removed upon application of a suitable source of radiation to give binding members, that are biotin or a functionally analogous compound having substantially the same binding affinity for avidin or avidin analogs as does biotin.

In another preferred embodiment, avidin or an avidin analog is incubated with activated binding members on the surface until the avidin binds strongly to the binding members. The avidin so immobilized on predefined regions of the surface can then be incubated with a desired receptor or conjugate of a desired receptor. The receptor will preferably be biotinylated, e.g., a biotinylated antibody, when avidin is immobilized on the predefined regions of the surface. Alternatively, a preferred embodiment will present an avidin/biotinylated receptor complex, which has been previously prepared, to activated binding members on the surface.

II. FINGERPRINTING

The above section on generation of reagents for sequencing provides specific reagents useful for fingerprinting applications. Fingerprinting embodiments may be applied towards polynucleotide fingerprinting, polypeptide fingerprinting, cell and tissue classification, cell and tissue temporal development stage classification, diagnostic tests, forensic uses for individual identification, classification of organisms, and genetic screening of individuals. Mapping applications are also described below.

A. Polynucleotide Fingerprint

Polynucleotide fingerprinting may use reagents similar to those described above for probing a sequence for the presence of specific subsequences found therein. Typically, the subsequences used for fingerprinting will be longer than the sequences used in oligonucleotide sequencing. In particular, specific long segments may be used to determine the similarity of different samples of nucleic acids. They may also be used to fingerprint whether specific combinations of information are provided therein. Particular probe sequences are selected and attached in a positional manner to a

substrate. The means for attachment may be either using a caged biotin method described, e.g., in U.S.S.N. 07/612,671 (caged biotin CIP), or by another method using targeting molecules. For example, a short polypeptide of specific sequence may be attached to an oligonucleotide and targeted to specific positions on a substrate having antibodies attached thereto, the antibodies exhibiting specificity for binding to those short peptide sequences. In another embodiment, an unnatural nucleotide or similar complementary binding molecule may be attached to the fingerprinting probe and the probe thereby directed towards complementary sequences on a VLSIPS substrate. Typically, unnatural nucleotides would be preferred, e.g., unnatural optical isomers, which would not interfere with natural nucleotide interactions.

Having produced a substrate with particular fingerprint probes attached thereto at positionally defined regions, the substrate may be used in a manner quite similar to the sequencing embodiment to provide information as to whether the fingerprint probes are detecting the corresponding sequence in a target sequence. This will often provide information similar to a Southern blot hybridization.

B. <u>Polypeptide Fingerprint</u>

A polypeptide fingerprint may be performed using antibodies which recognize specific antigens on the polypeptide. For example, monoclonal antibodies which recognize specific sequences or antigens on a polypeptide may be used to determine whether those epitopes are found on a particular protein. For example, particular patterns of epitopes would be found on various types of proteins. This will lead to the discovery that specific epitopes, or antigenic determinants, which are characteristic of, e.g., beta sheet segments, will be identified as will particular different types of domains in various protein types. Thus, a screening method may be devised which can classify polypeptides, either native or denatured, into various new classes defined by the epitopes existing thereon.

In addition, once the substrate is generated in the manners described above, a target peptide is exposed to the substrate. The target may be either native or denatured, though the conditions used to denature the polypeptide may interfere with the specific interaction between the polypeptide and the recognition reagent. This method is not dependent on the fact that the polypeptide is a single chain, thus protein complexes may also be fingerprinted using this methodology. Structures such as multi-subunit proteins, associations of proteins, ribosomes, nucleosomes, and other small cellular structures may also be fingerprinted and classified according to the presence of specific recognizable features thereon.

Peptide fingerprinting may be useful, for example, in correlating with particular physiological conditions or developmental stages of a cell or organism. Thus, a biological sample may be fingerprinted to determine the presence in that sample of a plurality of different polypeptides which are each individually fingerprinted. In an alternative embodiment, a polypeptide itself is not fingerprinted but a biological sample is fingerprinted searching for specific epitopes, e.g., polypeptide, carbohydrate, nucleic acid, or any of a number of other specific recognizable structural features.

The conditions for the interactions using antibodies is described, e.g., in Harlow and Lane (1988) Antibodies: A Laboratory Manual, Cold Spring Harbor Press, New York. The conditions should be titrated for temperature, buffer composition, time, and other important parameters in an antibody interaction.

C. Cell Classification Scheme

The present invention can be used for cell classification using fingerprinting type technology as described above in the polypeptide fingerprint. Classes of cells are typically defined by the presence of common functions which are usually reflected by structural features. Thus, a plant cell is classified differently from an animal cell by a number of structural features. Given an unknown cell, the present invention provides improved means for distinguishing

the different cell types. Once a cell classification scheme is developed and the structural features which define it are identified using the present invention, homogeneous cell population expressing these features may be separated from others. Standard cell sorters may be coupled with recognition reagents and labels which can distinguish various cell types.

a. T-Cell Classes

T-cell classes are defined on the basis of expression of particular antigens characteristic of each class. For example, mouse T-cell differentiation markers include the LY antigens. With the plurality of different antigens which may be tested using antibody or other recognition reagents, new populations and classes of cells may be defined. For example, different neural cell types may be defined on the basis of cell surface antigens. Different tissue types will be defined on the basis of tissue specific antigens. Developmental cell classes will be similarly defined. All of these screenings can make use of the VLSIPS substrates with specific recognition molecules attached thereto. The substrates are exposed to the cell types directly, assaying for attachment of cells to specific regions, or are exposed to products of a population of cells, e.g., a supernatant, or a cell lysate.

Once a cell classification scheme has been correlated with specific structural markers therein, reagents which recognize those features may be developed and used in a fluorescence activated cell sorter as described, e.g., in Dangl, J. and Herzenberg (1982) J. Immunological Methods 52: 1-14; and Becton Dickinson, Fluorescence Activated Cell Sorters Division, San Jose, California. This will provide a homogeneous population of cells whose function has been defined by structure.

b. B-Cell Classes

The present cell classification scheme may also be used to determine specific B-cell classes. For example, B-cells specific for producing IgM, IgG, IgD, IgE, and IgA may be defined by the internal expression of specific mRNA sequences

encoding each type of immunoglobulin. The classification scheme may depend on either extracellularly expressed markers which are correlated as being diagnostic of specific stages in development, or intracellular mRNA sequences which indicate particular functions.

D. Temporal Development Scheme

1. Developmental Antigens

The present fingerprinting invention also allows cell classification by expression of developmental antigens. For example, a lymphocyte stem cell expresses a particular combination of antigens. As the lymphocyte develops through a program developmental scheme, at various stages it expresses particular antigens which are diagnostic of particular stages in development. Again, the fingerprinting methodology allows for the definition of specific structural features which are diagnostic of developmental or functional features which will allow classification of cells into temporal developmental classes. Cells, products of those cells, or lysates of those cells will be assayed to determine the developmental stage of the source cells. In this manner, once a developmental stage is defined, specific synchronized populations of cells will be selected out of another population. These synchronized populations may be very important in determining the biological mechanisms of development.

2. Developmental mRNA Expression

Besides expressed antigens, the present invention also allows for fingerprinting of the mRNA population of a cell. In this fashion, the mRNA population, which should be a good determinant of developmental stage, will be correlated with other structural features of the cell. In this manner, cells at specific developmental stages will be characterized by the intracellular environment, as well as the extracellular environment. The present invention also allows the combination of definitions based, in part, upon antigens and, in part, upon mRNA expression.

In one embodiment, the two may be combined in a single incubation step. A particular incubation condition may be found which is compatible with both hybridization recognition non-hybridization recognition molecules. Thus, e.g., an incubation condition may be selected which allows both specificity of antibody binding and specificity of nucleic acid hybridization. This allows simultaneous performance of both types of interactions on a single matrix. Again, where developmental mRNA patterns are correlated with structural features, or with probes which are able to hybridize to intracellular mRNA populations, a cell sorter may be used to sort specifically those cells having desired mRNA population patterns.

E. Diagnostic Tests

The present invention also provides the ability to perform diagnostic tests. Diagnostic tests typically are based upon a fingerprint type assay, which tests for the presence of specific diagnostic structural features. Thus, the present invention provides means for viral strain identification, bacterial strain identification, and other diagnostic tests using positionally defined specific reagents. The present invention also allows for determining a spectrum of allergies, diagnosing a biological sample for any or all of the above, and testing for many other conditions.

1. Viral Identification

The present invention provides reagents and methodology for identifying viral strains. The specific reagents may be either antibodies or recognition proteins which bind to specific viral epitopes preferably surface exposed, but may make use of internal epitopes, e.g., in a denatured viral sample. In an alternative embodiment, the viral genome may be probed for specific sequences which are characteristic of particular viral strains. As above, a combination of the two may be performed simultaneously in a single interaction step, or in separate tests, e.g., for both genetic characteristics and epitope characteristics.

2. Bacterial Identification

Similar techniques will be applicable to identifying a bacterial source. This may be useful in diagnosing bacterial infections, or in classifying sources of particular bacterial species. For example, the bacterial assay may be useful in determining the natural range of survivability of particular strains of bacteria across regions of the country or in different ecological niches.

3. Other Microbiological Identifications

The present invention provides means for diagnosis of other microbiological and other species, e.g., protozoal species and parasitic species in a biological sample, but also provides the means for assaying a combination of different infections. For example, a biological specimen may be assayed for the presence of any or all of these microbiological species. In human diagnostic uses, typical samples will be blood, sputum, stool, urine, or other samples.

4. Allergy Tests

An immobilized set of antigens may be attached to a solid substrate and, instead of the standard skin reaction tests, a blood sample may be assayed on such a substrate to determine the presence of antibodies, e.g., IgE or other type antibodies, which may be diagnostic of an allergic or immunological susceptibility. A standard radioallergosorbent test (RAST) may be used to check a much larger population of antigens.

In addition, an allergy like test may be used to diagnose the immunological history of a particular individual. For example, by testing the circulating antibodies in a blood sample, which reflects the immunological history and memory of an individual, it may be determined what infections may not have been historically presented to the immune system. In this manner, it may be possible to specifically supplement an immune system for a short period of time with IgG fractions made up of specific types of gamma globulins. Thus, hepatitis gamma

globulin injections may be better designed for a particular environment which a person is expected to be exposed. This also provides the ability to identify genetically equivalent individuals who have immunologically different experiences. Thus, a blood sample from an individual who has a particular combination of circulating antibodies will likely be different from the combination of circulating antibodies found in a genetically similar or identical individual. This could allow for the distinction between clones of particular animals, e.g., mice, rats, or other animals.

F. Individual Identification

The present invention provides the ability to fingerprint and identify a genetic individual. This individual may be a bacterial or lower microorganism, as described above in diagnostic tests, or of a plant or animal. An individual may be identified genetically or immunologically, as described.

1. Genetic

Genetic fingerprinting has been utilized in comparing different related species in Southern hybridization blots. Genetic fingerprinting has also been used in forensic studies, see, e.g., Morris et al. (1989) J. Forensic Science 34: 1311-1317, and references cited therein. As described above, an individual may be identified genetically by a sufficiently large number of probes. The likelihood that another individual would have an identical pattern over a sufficiently large number of probes may be statistically negligible. However, it is often quite important that a large number of probes be used where the statistical probability of matching is desired to be particularly low. In fact, the probes will optimally be selected for having high heterogeneity among the population. In addition, the fingerprint method may make use of the pattern of homologies indicated by a series of more and more stringent washes. Then, each position has both a sequence specificity and a homology measurement, the combination of which greatly increases the number of dimensions and the statistical

likelihood of a perfect pattern match with another genetic individual.

2. Immunological

As indicated above in the diagnostic tests, it is possible to identify a particular immune system within a genetically homogeneous class of organisms by virtue of her immunological history. For example, a large colony of cloned mice may be distinguishable by virtue of each immunological history. For example, one mouse may have had an immunological response to exposure to antigen A to which her genetically identical sibling may have not been exposed. By virtue of this differential history, the first of the pair will likely have a high antibody titer against the antigen A whereas her genetically identical sibling will have not had a response to that antigen by virtue of never having been exposed to it. For this reason, immune systems may be identified by their immunological memories. Thus, immunological experience may also be a means for identifying a particular individual at a particular moment in her lifetime.

This same immunological screening may be used for other sorts of identifiable biological products. For example, an individual may be identified by her combination of expressed proteins. These proteins may reflect a physiological state of the individual, and would thus be useful in certain circumstances where diagnostic tests may be performed. For example, an individual may be identified, in part, by the presence of particular metabolic products.

In fact, a plant origin may be determined by virtue of having within its genome an unnatural sequence introduced to it by genetic breeders. Thus, a marker nucleic acid sequence may be introduced as a means to determine whether a genetic strain of a plant or animal originated from another particular source.

G. Genetic Screening

1. test alleles with markers

The present invention provides for the ability to screen for genetic variations of individuals. For example, a number of genetic diseases are linked with specific alleles. See, e.g., Scriber, C. et al. (eds.) (1989) The Metabolic Bases of Inherited Disease, McGraw-Hill, New York. In one embodiment, cystic fibrosis has been correlated with a specific gene, see, Gregory et al. (1990) Nature 347: 382-386. A number of alleles are correlated with specific genetic deficiencies. See, e.g., McKusick, V. (1990) Genetic Inheritance in Man: Catalogs of Autosomal Dominant, Autosomal Recessive, and X-linked Phenotypes, Johns Hopkins University Press, Baltimore; Ott, J. (1985) Analysis of Human Genetic Linkage, Johns Hopkins University Press, Baltimore; Track, R. et al. (1989) Banbury Report 32: DNA Technology and Forensic Science, Cold Spring Harbor Press, New York; each of which is hereby incorporated herein by reference.

2. Amniocentesis

Typically, amniocentesis is used to determine whether chromosome translocations have occurred. The mapping procedure may provide the means for determining whether these translocations have occurred, and for detecting particular alleles of various markers.

III. MAPPING

A. Positionally Located Clones

The present invention allows for the positional location of specific clones useful for mapping. For example, caged biotin may be used for specifically positioning a probe to a location on a matrix pattern.

In addition, the specific probes may be positionally directed to specific locations on a substrate by targeting. For example, polypeptide specific recognition reagents may be attached to oligonucleotide sequences which can be complementarily targeted to specific locations on a VLSIPS substrate. Hybridization conditions, as applied for

oligonucleotide probes, will be used to target the reagents to locations on a substrate having complementary oligonucleotides synthesized thereon. In another embodiment, oligonucleotide probes may be attached to specific polypeptide targeting reagents such as an antigen or antibody. These reagents can be directed towards a complementary antigen or antibody already attached to a VLSIPS substrate.

In another embodiment, an unnatural nucleotide which does not interfere with natural nucleotide complementary hybridization may be used to target oligonucleotides to particular positions on a substrate. Unnatural optical isomers of natural nucleotides should be ideal candidates.

In this way, short probes may be used to determine the mapping of long targets or long targets may be used to map the position of shorter probes. See, e.g., Craig et al. 1990 Nuc. Acids Res. 18: 2653-2660.

B. Positionally Defined Clones

Positionally defined clones may be transferred to a new substrate by either physical transfer or by synthetic means. Synthetic means may involve either a production of the probe on the substrate using the VLSIPS synthetic methods, or may involve the attachment of a targeting sequence made by VLSIPS synthetic methods which will target that positionally defined clone to a position on a new substrate. Both methods will provide a substrate having a number of positionally defined probes useful in mapping.

IX. Conclusion

The present inventions provide greatly improved methods and apparatus for synthesis of polymers on substrates. It is to be understood that the above description is intended to be illustrative and not restrictive. Many embodiments will be apparent to those of skill in the art upon reviewing the above description. By way of example, the invention has been described primarily with reference to the use of photoremovable protective groups, but it will be readily recognized by those of skill in the art that sources of radiation other than light

could also be used. For example, in some embodiments it may be desirable to use protective groups which are sensitive to electron beam irradiation, x-ray irradiation, in combination with electron beam lithograph, or x-ray lithography techniques. Alternatively, the group could be removed by exposure to an electric current. The scope of the invention should, therefore, be determined not with reference to the above description, but should instead be determined with reference to the appended claims, along with the full scope of equivalents to which such claims are entitled.

All publications and patent applications referred to herein are incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually incorporated by reference. The present invention now being fully described, it will be apparent to one of ordinary skill in the art that many changes and modifications can be made thereto without departing from the spirit or scope of the appended claims.

WHAT IS CLAIMED IS:

1. A composition comprising a plurality of positionally distinguishable sequence specific reagents attached to a solid substrate, which reagents are capable of specifically binding to a predetermined subunit sequence of a preselected multi-subunit length having at least three subunits, said reagents representing substantially all possible sequences of said preselected length.

2. A composition of Claim 1, wherein said subunit sequence is a polynucleotide or a polypeptide.

3. A composition of Claim 1, wherein said preselected multi-subunit length is five subunits and said subunit sequence is a polynucleotide sequence.

4. A composition of Claim 1, wherein said specific reagent is an oligonucleotide of at least about five nucleotides.

5. A composition of Claim 1, wherein said specific reagent is a monoclonal antibody.

6. A composition of Claim 1, wherein said specific reagents are all attached to a single solid substrate.

7. A composition of Claim 1, wherein said reagents comprise about 3000 different sequences.

8. A composition of Claim 1, wherein said reagents represents at least about 25% of the possible subsequences of said preselected length.

9. A composition of Claim 1, wherein said reagents are localized in regions of the substrate having a density of at least 25 regions per square centimeter.

10. A composition of Claim 6, wherein said substrate has a surface area of less than about 4 square centimeters.

11. A method of analyzing a sequence of a polynucleotide or a polypeptide, said method comprising the step of:

a) exposing said polynucleotide or polypeptide to a composition of Claim 1.

12. A method of identifying or comparing a target sequence with a reference, said method comprising the step of:

a) exposing said target sequence to a composition of Claim 1;

b) determining the pattern of positions of said reagents which specifically interact with said target sequence; and

c) comparing said pattern with the pattern exhibited by said reference when exposed to said composition.

13. A method for sequencing a segment of a polynucleotide comprising the steps of:

a) combining:

i) a substrate comprising a plurality of chemically synthesized and positionally distinguishable oligonucleotides capable of recognizing defined oligonucleotide sequences; and

ii) a target polynucleotide; thereby forming high fidelity matched duplex structures of complementary subsequences of known sequence; and

b) determining which of said reagents have specifically interacted with subsequences in said target polynucleotide.

14. A method of Claim 13, wherein said segment is substantially the entire length of said polynucleotide.

15. A method for sequencing a polymer, said method comprising the steps of:

a) preparing a plurality of reagents which each specifically bind to a subsequence of preselected length;

b) positionally attaching each of said reagents to one or more solid phase substrates, thereby producing substrates of

positionally definable sequence specific probes;

c) combining said substrates with a target polymer whose sequence is to be determined; and

d) determining which of said reagents have specifically interacted with subsequences in said target polymer.

16. A method of Claim 15, wherein said substrates are beads.

17. A method of Claim 15, wherein said plurality of reagents comprise substantially all possible subsequences of said preselected length found in said target.

18. A method of Claim 15, wherein said solid phase substrates are a single substrate having attached thereto reagents recognizing substantially all possible subsequences of preselected length found in said target.

19. A method of Claim 15, further comprising the step of analyzing a plurality of said recognized subsequences to assemble a sequence of said target polymer.

20. A method of Claim 16, wherein at least some of said plurality of substrates have one subsequence specific reagent attached thereto, and said substrates are coded to indicate the specificity of said reagent.

21. A method of using a fluorescent nucleotide to detect interactions with oligonucleotide probes of known sequence, said method comprising:

a) attaching said nucleotide to a target unknown polynucleotide sequence, and

b) exposing said target polynucleotide sequence to a collection of positionally defined oligonucleotide probes of known sequences to determine the sequences of said probes which interact with said target.

22. A method of Claim 21, further comprising the step of:

a) collating said known sequences to determine the overlaps of said known sequences to determine the sequence of said target sequence.

23. A method of mapping a plurality of sequences relative to one another, said method comprising:

a) preparing a substrate having a plurality of positionally attached sequence specific probes are attached;

b) exposing each of said sequences to said substrate, thereby determining the patterns of interaction between said sequence specific probes and said sequences; and

c) determining the relative locations of said sequence specific probe interactions on said sequences to determine the overlaps and order of said sequences.

24. A method of Claim 23, wherein said sequence specific probes are oligonucleotides.

25. A method of Claim 23, wherein said sequences are nucleic acid sequences.

SEQUENCING BY HYBRIDIZATION OF A TARGET NUCLEIC ACID TO A MATRIX OF DEFINED OLIGONUCLEOTIDES

ABSTRACT

The present invention provides methods and apparatus for sequencing, fingerprinting and mapping biological macromolecules, typically biological polymers. The methods make use of a plurality of sequence specific recognition reagents which can also be used for classification of biological samples, and to characterize their sources.

This application is a continuation-in-part application of commonly assigned patent applications Pirrung et al., U.S.S.N. 07/362,901 (VLSIPS parent) filed on June 7 1989; and Pirrung et al., U.S.S.N. 07/492,462 (VLSIPS CIP), filed on March 7, 1990, which are hereby incorporated herein by reference. Additional commonly assigned applications Barrett et al., U.S.S.N. 07/435,316 (caged biotin parent) filed November 13, 1989; and Barrett et al., U.S.S.N. 07/612,671 (caged biotin CIP), filed November 13, 1990 are also incorporated herein by reference. Additional applications Pirrung et al., U.S.S.N. __/___,___, attorney docket number 11509-28 (automated VLSIPS); and Dower et al., U.S.S.N. __/___,___, attorney docket number 11509-26 (microfluorescence sequencing), which are also commonly assigned and filed on the same day as this application, are also hereby incorporated herein by reference.

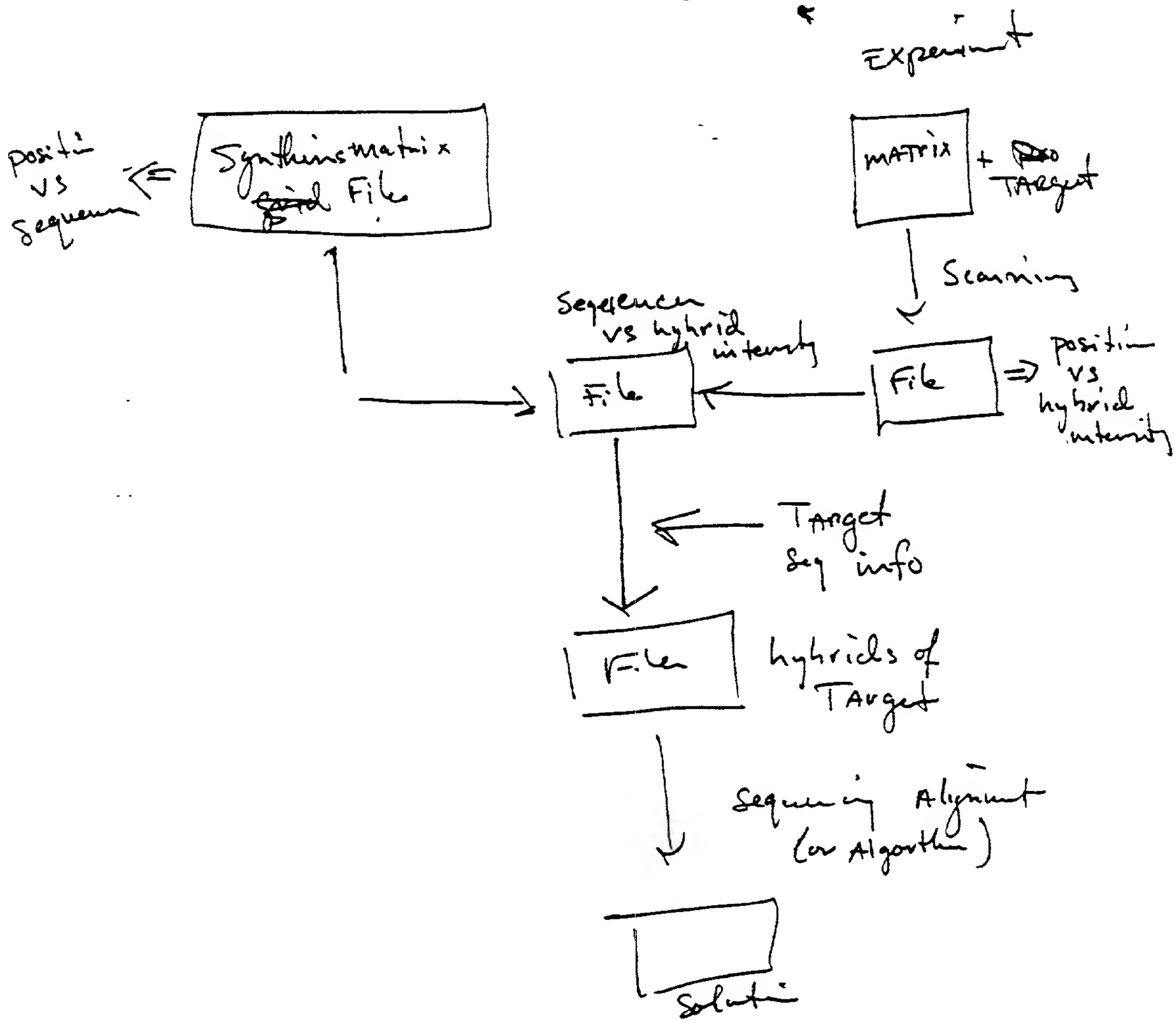

Experiment
Synthesis matrix File
positive vs sequence
MATRIX
+ TARGET
Scanning
Sequence vs hybrid intensity
File
File
positive vs hybrid intensity
Target Seq info
File
hybrids of TArget
Sequencing Alignment (or Algorithm)
Solution

Figure 1

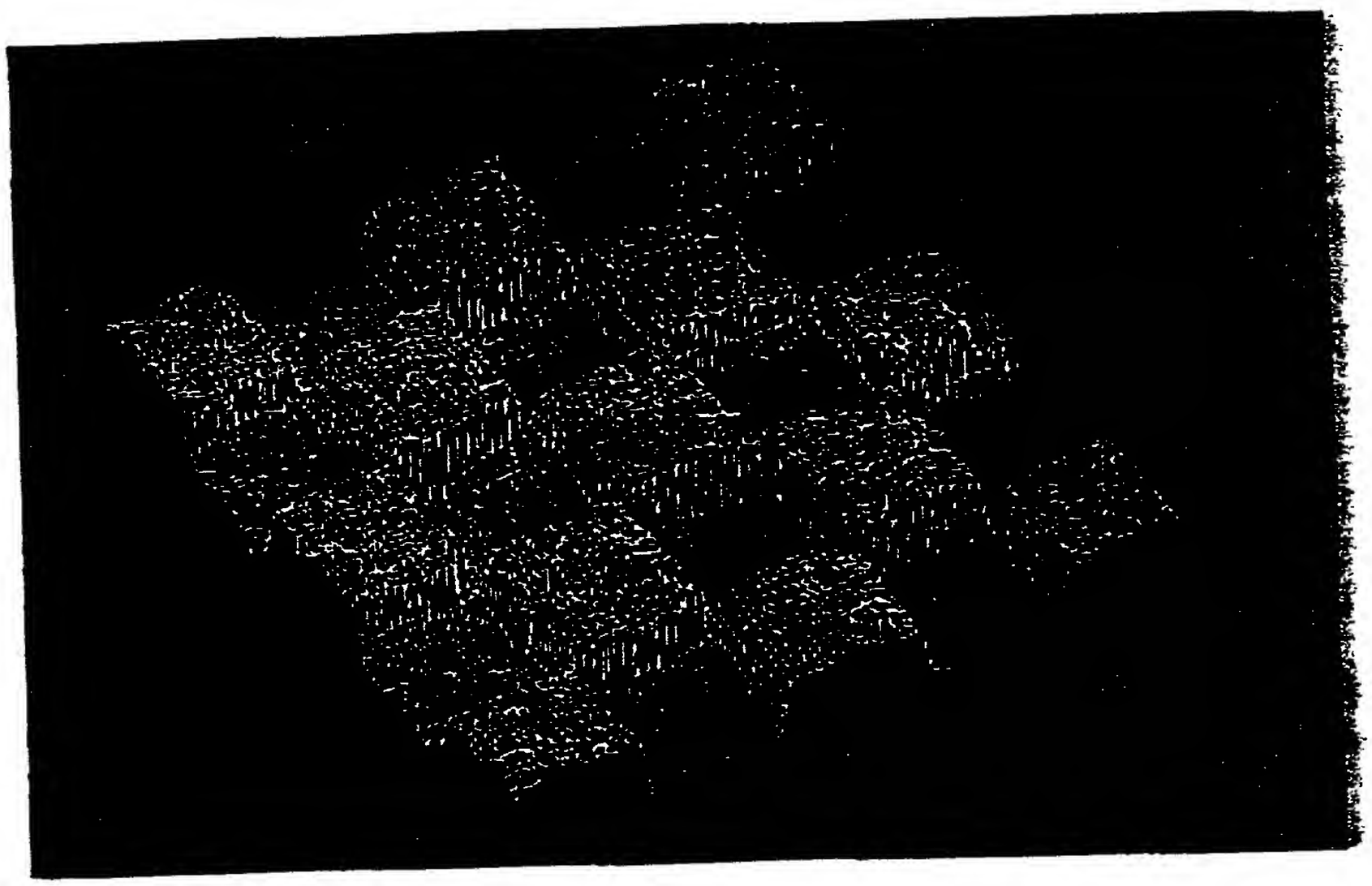


Figure 2

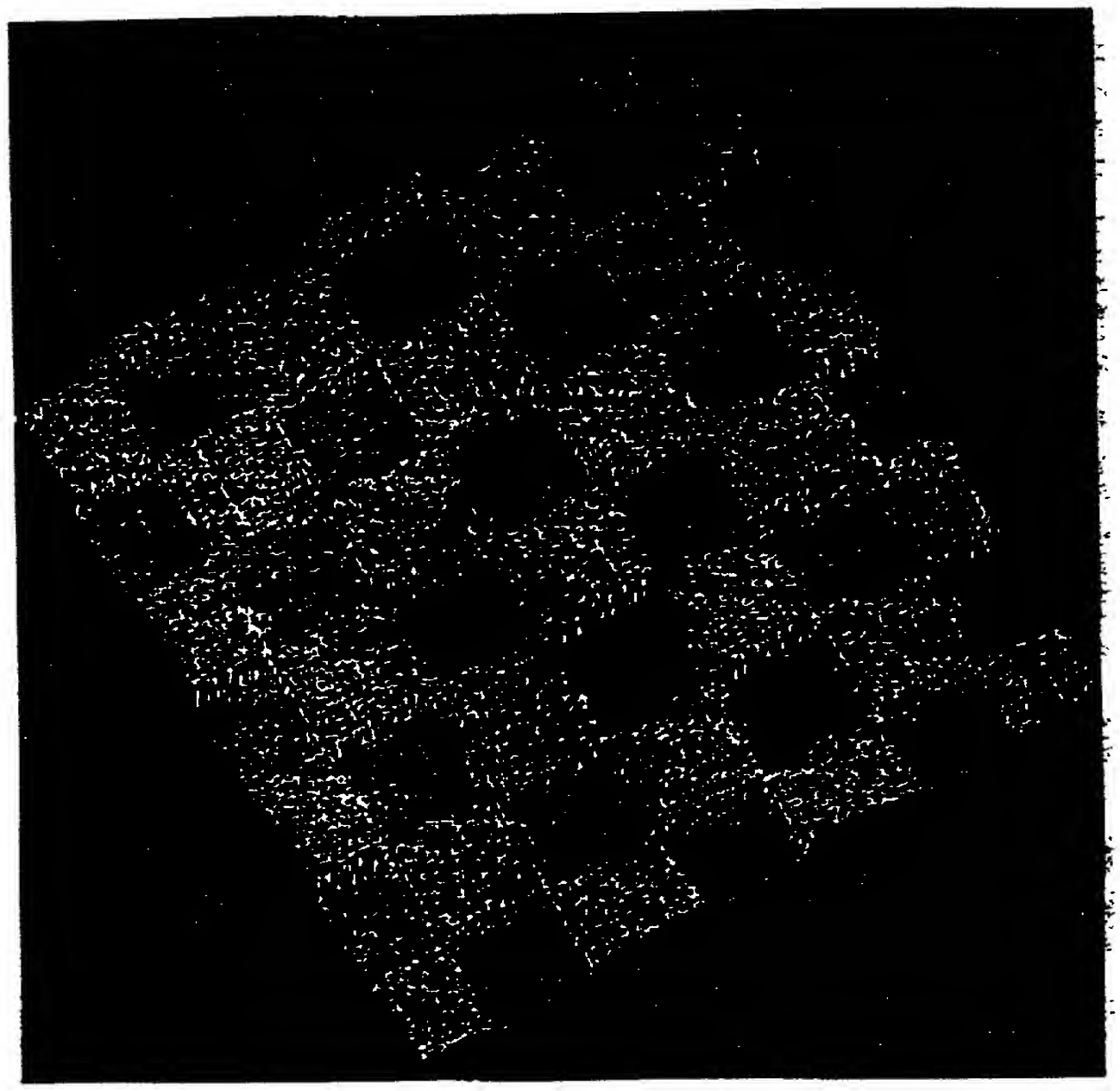


Figure 3

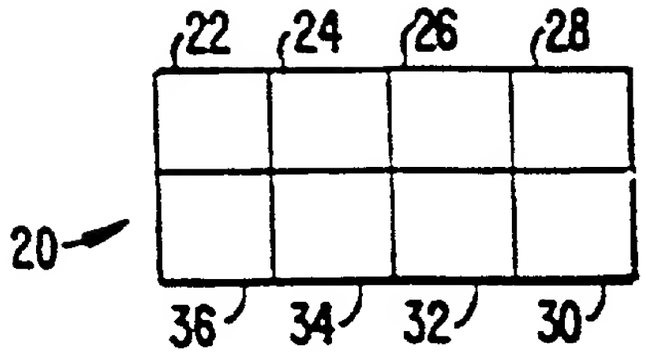


FIG. 4 A.

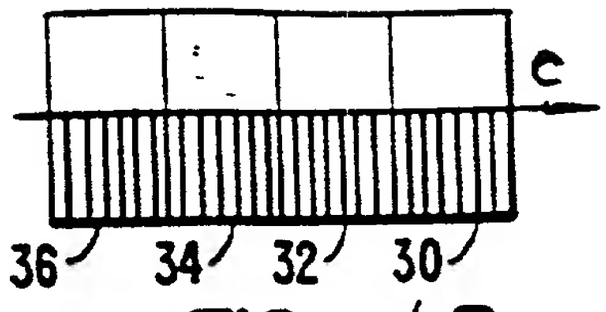


FIG. 4 B.

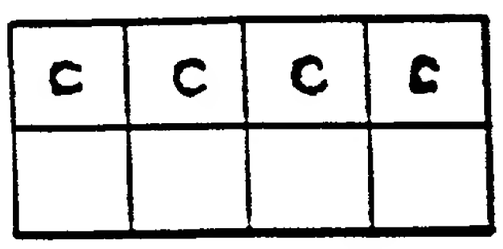


FIG. 4 C.

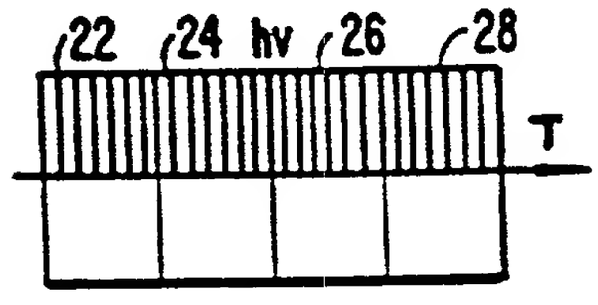


FIG. 4 D.

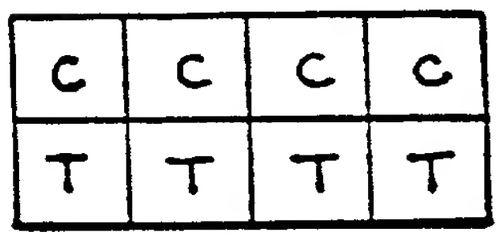


FIG. 4 E.

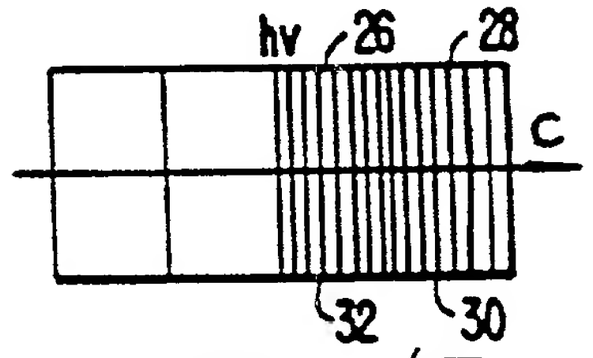


FIG. 4 F.

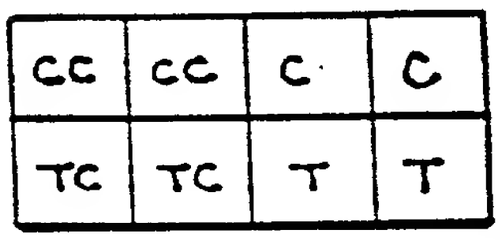


FIG. 4 G.

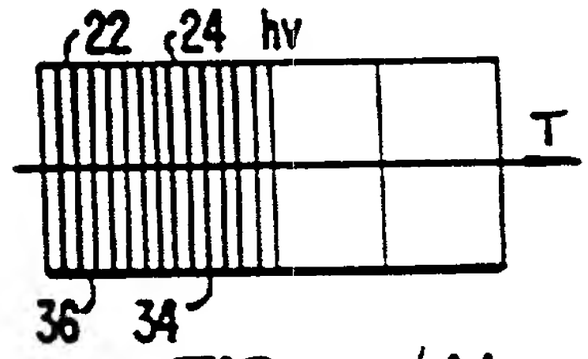


FIG. 4 H.

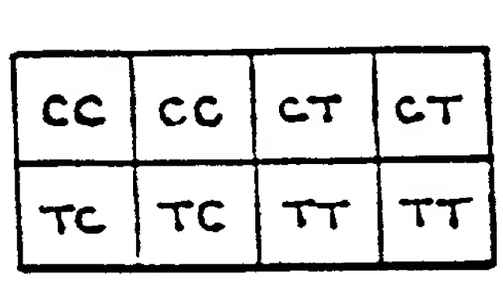


FIG. 4 I.

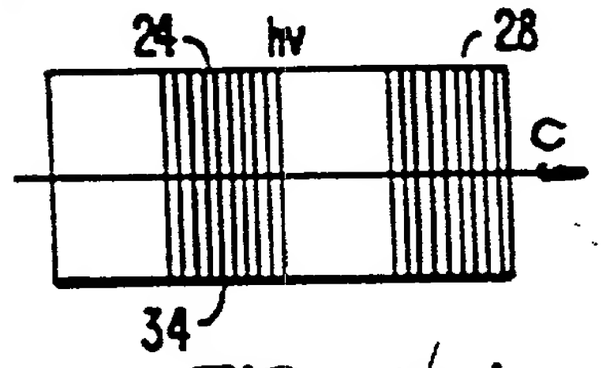


FIG. 4 J.

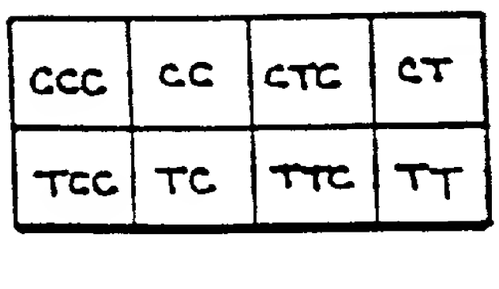


FIG. 4 K.

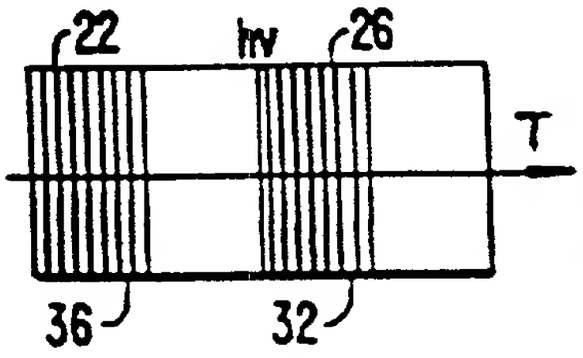


FIG. 4 L.

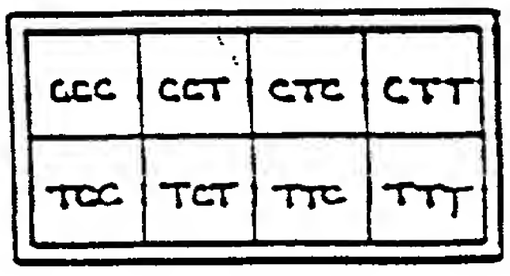


FIG. 4 M.

Figure 4 A through M

11509-30
ATTORNEY DOCKET NO.

F

DECLARATION AND POWER OF ATTORNEY

As a below named inventor, I hereby declare that:
My residence, post office address and citizenship are as stated below next to my name; I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural inventors are named below) of the subject matter which is claimed and for which a patent is sought on the invention entitled: SEQUENCING BY HYBRIDIZATION OF A TARGET NUCLEIC ACID TO A MATRIX OF DEFINED OLIGONUCLEOTIDES

the specification of which ☐ is attached hereto or ☒ was filed on December 6, 1990 as Application Serial No 07/624,114 and was amended on _____ (if applicable).

I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment referred to above. I acknowledge the duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, §1.56(a). I claim foreign priority benefits under Title 35, United States Code, §119 of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed:

Prior Foreign Application(s)

COUNTRY	APPLICATION NUMBER	DATE OF FILING	PRIORITY CLAIMED UNDER 35 U.S.C. 119
			Yes____ No____
			Yes____ No____

I claim the benefit under Title 35, United States Code, §120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, §112, I acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, §1.56(a) which occurred between the filing date of the prior application and the national or PCT international filing date of this application:

APPLICATION SERIAL NO.	DATE OF FILING	STATUS
492,462	March 7, 1990	☐ Patented ☒ Pending ☐ Abandoned
362 991	June 7, 1989	☐ Patented ☒ Pending ☐ Abandoned

POWER OF ATTORNEY: As a named inventor, I hereby appoint the following attorney(s) and/or agent(s) who are partners and associates in the firm of Townsend and Townsend to prosecute this application and transact all business in the Patent and Trademark Office connected therewith. Edwin P. Ching, Reg. No. 34,090
William M. Smith, Reg. No. 30,223

SEND CORRESPONDENCE TO:	DIRECT TELEPHONE CALLS TO: *(name, registration number, and telephone number)*
TOWNSEND and TOWNSEND Steuart Street Tower, One Market Plaza San Francisco, CA 94105	Edwin P. Ching, Reg. No. 34,090 ☐ (415) 543-9600 or ☒ (415) 326-2400

201	FULL NAME OF INVENTOR	Last Name: FODOR	First Name: STEPHEN		Middle Name or Initial: P.A.
	RESIDENCE & CITIZENSHIP	City: PALO ALTO	State or Foreign Country: CALIFORNIA		Country of Citizenship: U.S.A.
	POST OFFICE ADDRESS	Post Office Address: 3863 Nathan Way	City: Palo Alto	State or Country: California	Zip Code: 94303
202	FULL NAME OF INVENTOR	Last Name: SOLAS	First Name: DENNIS		Middle Name or Initial: W.
	RESIDENCE & CITIZENSHIP	City: SAN FRANCISCO	State or Foreign Country: CALIFORNIA		Country of Citizenship: U.S.A.
	POST OFFICE ADDRESS	Post Office Address: 50 Gardenside Dr., #13	City: San Francisco	State or Country: California	Zip Code: 94131
203	FULL NAME OF INVENTOR	Last Name: DOWER	First Name: WILLIAM		Middle Name or Initial: J.
	RESIDENCE & CITIZENSHIP	City: MENLO PARK	State or Foreign Country: CALIFORNIA		Country of Citizenship: U.S.A.
	POST OFFICE ADDRESS	Post Office Address: 761 Partridge Avenue	City: Menlo Park	State or Country: California	Zip Code: 94025

I further declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Signature of Inventor 201	Signature of Inventor 202	Signature of Inventor 203
Stephen Fodor	Dennis Solas	William J. Dower
Date: Jan 28, 1991	Date: 1/24/91	Date: FEB 1, 1991